

**STIC-Biotech/ChemLib**

1060288

**From:** Fredman, Jeffrey  
**Sent:** Monday, October 20, 2003 4:02 PM  
**To:** STIC-Biotech/ChemLib  
**Cc:** Leffers, Gerald  
**Subject:** FW: 10/029,137

PLEASE RUSH.

I Approve.

Jeff Fredman

(To searcher: Gerald needs the regular search for SEQ ID NO: 1, 2, 7-10, as well as an oligomer search of SEQ ID NO: 1 and a nucleic acid search of SEQ ID NO: 2 along with the protein search)  
(Gerry, if you want anything else, or different, let STIC know)

-----Original Message-----

**Fr m:** Leffers, Gerald  
**Sent:** Monday, October 20, 2003 3:04 PM  
**To:** Fredman, Jeffrey  
**Subject:** 10/029,137

Hi Jeff,

please approve a RUSH search for SEQ ID NOS: 1 & 2 for this application (~1.3 kb and 180 amino acids). Please also search SEQ ID NOS: 7-10 (primers of 18-20 nucleotides).

Claims drawn to (a) nucleic acids that hybridize to SEQ ID NO: 1 under stringent conditions,

(b) a nucleic acid that encodes SEQ ID NO: 2,

(c) a nucleic acid amplified by primers 7 and 8 (SEQ ID NOS: 7 and 8),

(d) a nucleic acid amplified by primers 9-10 (seq id nos 9-10),

(e) nucleic acid with 90% identity to the coding region of SEQ ID No: 1, and

(f) a pair of primers that can amplify a nucleic acid encoding a human EG-1 polypeptide sequence (i.e. SEQ ID NO: 2).

Thanks for your help Jeff, Gerry

*Gerald G. Leffers Jr., PhD*  
Examiner, Art Unit 1636  
Crystal Mall 1, Room 11A09  
703-308-6232

STIC-Biotech/ChemLib  
Biology & Chemical Library  
Tel: 703-308-4498  
gndelval@uspl.gov

Searcher: Jan  
Phone: 4438  
Location: \_\_\_\_\_  
Date Picked Up: 10/20  
Date Completed: 10/21  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: 30  
Online time: 20

TYPE OF SEARCH: ☒  
NA Sequences: ☒  
AA Sequences: ☒  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: ☒  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_





[illegible][illegible]







DEFINITION Homo sapiens, hypothetical protein DKF2p434n195, clone MGC:20562 IMAGE:5936713, mRNA, complete cds.

ACCESSION BC011936

VERSION 1

KEYWORDS MGC

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1083)

AUTHORS Strausberg, R.

TITLE Direct Substitution

JOURNAL Submitted (30-JUL-2001) National Institutes of Health, Mammalian Genome Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A93, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>

COMMENT Contact: MGC help desk

Email: [gcgaps-ramail.nih.gov](mailto:gcgaps-ramail.nih.gov)

Tissue Procurement: ECTD/BFP

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILN)

DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC)

Galbraithsburg, Maryland

Web site: <http://www.nisc.nih.gov/>

Contact: [nisc\\_mgc@nih.gov](mailto:nisc_mgc@nih.gov)

Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Stinkley, C., Brooks, S., Dietrich, N.L., Guan, J., Gupta, J., Ho, S.L., Karlins, E., Legaspi, R., Lin, X., Maduro, Q.B., Masello, C., Mastlian, S.D., McLoskey, J.C., McDowell, J., Pearson, R., Snyder, B., Stambrope, S., Thomas, P., Thompson, E., Touchman, J., Tsukaguchi, C., Vogel, J., Walker, M.A., Zhang, H., and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILN at: <http://image.llnl.gov>

Series: IMAGE Plate: 2d Row: 1 Column: 4

This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency CRF analysis.

Location/Qualifiers

1..1083

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="ccs:ID:80306"

/db\_xref="taxon:9606"

/cds="MGC:20562 IMAGE:5936713"

/color="MGC:20562 IMAGE:5936713"

/tissue\_type="Lung, small cell carcinoma"

/clone\_lib="NIH MGC 7"

/lab\_host="DH10B-R"

/note="Vector: pOTB7"

14..1550

/codon\_start=1

/product="hypothetical protein DKF2p434n195"

/protein\_id="AA11936.1"

/db\_xref="GI:15080367"

/translation="MAAPLQMGSPGQHPGAPGLPGASLLQAPGAPRPSSTLVCLESSEAPGASLVQVYNGVLCQSRHTGVCQKXKFIHQVCECFQRRQLQVQKPEQKEDVSELNFQKVALVQHLTKLRHQVQVLEIDVQKPKADIPQGS LAYLEQASAN:PAFLPKP"

BASE COUNT 294 a 216 c 259 g 314 t

ORIGIN

Query Match 82.2% Score 1053.4 DB 9; Length 1083;

Best Local Similarity 99.9% Pred. No. 6.7e-238;

Matches 1054; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 70 CCGGCGGCGGCTTGGGAAAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 129

QY 122 TTCCAGCAGTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 141

DB 140 TTCCAGCAGTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 189

QY 182 CTGAGTCGAGTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 241

DB 240 CTGAGTCGAGTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 249

QY 242 CTGAGTCGAGTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 301

DB 300 CTGAGTCGAGTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 309

QY 302 ATTTAGTTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 361

DB 360 ATTTAGTTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 369

QY 362 ATTTAGTTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 421

DB 420 ATTTAGTTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 429

QY 422 ATTTAGTTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 481

DB 480 ATTTAGTTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 489

QY 482 ATTTAGTTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 541

DB 540 ATTTAGTTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 549

QY 542 ATTTAGTTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 601

DB 600 ATTTAGTTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 609

QY 602 ATTTAGTTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 661

DB 660 ATTTAGTTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 669

QY 662 ATTTAGTTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 721

DB 720 ATTTAGTTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 729

QY 722 ATTTAGTTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 781

DB 780 ATTTAGTTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 789

QY 782 ATTTAGTTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 841

DB 840 ATTTAGTTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 849

QY 842 ATTTAGTTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 901

DB 900 ATTTAGTTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 909

QY 902 ATTTAGTTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 961

DB 960 ATTTAGTTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 969

QY 962 ATTTAGTTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 1021

DB 1020 ATTTAGTTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 1029

QY 1022 ATTTAGTTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 1081

DB 1080 ATTTAGTTATTTCTGGAGAGTATTTCTTACGAGCTCCAGCGATCCAGAAAT 1089

RESULT 7

AC006160 1793.0 bp DNA linear PRI 04-AUG 1999

LOCUS Homo sapiens chromosome 4 clone 2049.F14 map 4p16, complete

DEFINITION sequence









Source	1..12867	12867 bp	DNA	Linear	PR1 04-APR 2001
LOCUS	HS694B14/c				
DEFINITION	Human DNA sequence from clone RP4-694B14 on chromosome 20p11.1-11.22. Contains a novel KRAS box protein with 13 CH2 type zinc finger domains, a novel haloacid dehalogenase-like hydrolase family protein similar to (archaea) bacterial proteins, two putative novel genes, a novel pseudogene, ESTs, an STS, GSSE and three CpG islands. Complete sequence.				
ACCESSION	AL011673				
VERSION	AL011673.19	GI:11958366			
KEYWORDS	HTG; CpG island; haloacid dehalogenase; hydrolase; KRAS box; zinc finger.				
SOURCE	Homo sapiens				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryotic; Vertebrate; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Smith M.				
TITLE	Direct Submission				
JOURNAL	Submitted (25-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail: enquiries@hmpuquery@sanger.ac.uk				
COMMENT	On Dec 22, 2000 this sequence version replaced GI:11958366. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: EMBL, EMBL; SWISS, SWISSPROT; TrEMBL, TrEMBL; WPI, WORMPEP; Information on the WORMPEP database can be found at: <a href="http://www.sanger.ac.uk/Projects/C_elegans/wormpep">http://www.sanger.ac.uk/Projects/C_elegans/wormpep</a> . This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre (chromosome 20 Mapping Group). Further information can be found at: <a href="http://www.sanger.ac.uk/HGP/Chr20">http://www.sanger.ac.uk/HGP/Chr20</a> . RP4-694B14 is from the library RP41-4 constructed by the group of Pieter de Jong. For further details see <a href="http://www.chori.org/hacpac/home.htm">http://www.chori.org/hacpac/home.htm</a> . VECTR: PYPAC2				
FEATURES	This sequence is the entire insert of clone RP4-694B14. The true left end of clone RP41-40-NB is at 116161 in this sequence. The true right end of clone RP41-9616 is at 20586 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., paired quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one W13 subclone; and the assembly was confirmed by restriction digest.				
Source	Location/Qualifiers				
	1..12867				
	/organism="Homo sapiens"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:9606"				
	/chromosome="20"				
	/map="p11.1-11.22"				
	/clone="RP4-694B14"				
	/clone_lib="RP41-4"				
	1..74				
repeat_region	/note="AluJb repeat: matches 194..266 of consensus"				
repeat_region	787..1088				
repeat_region	/note="AluSp repeat: matches 1..303 of consensus"				
repeat_region	1513..1786				
repeat_region	/note="AluSc repeat: matches 1..295 of consensus"				
repeat_region	1818..2372				
repeat_region	/note="11MC3 repeat: matches 7230..7739 of consensus"				
repeat_region	2171..2669				
repeat_region	/note="AluJb repeat: matches 1..298 of consensus"				
repeat_region	2670..3427				

Source	1..12867	12867 bp	DNA	Linear	PR1 04-APR 2001
LOCUS	HS694B14/c				
DEFINITION	Human DNA sequence from clone RP4-694B14 on chromosome 20p11.1-11.22. Contains a novel KRAS box protein with 13 CH2 type zinc finger domains, a novel haloacid dehalogenase-like hydrolase family protein similar to (archaea) bacterial proteins, two putative novel genes, a novel pseudogene, ESTs, an STS, GSSE and three CpG islands. Complete sequence.				
ACCESSION	AL011673				
VERSION	AL011673.19	GI:11958366			
KEYWORDS	HTG; CpG island; haloacid dehalogenase; hydrolase; KRAS box; zinc finger.				
SOURCE	Homo sapiens				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryotic; Vertebrate; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Smith M.				
TITLE	Direct Submission				
JOURNAL	Submitted (25-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail: enquiries@hmpuquery@sanger.ac.uk				
COMMENT	On Dec 22, 2000 this sequence version replaced GI:11958366. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: EMBL, EMBL; SWISS, SWISSPROT; TrEMBL, TrEMBL; WPI, WORMPEP; Information on the WORMPEP database can be found at: <a href="http://www.sanger.ac.uk/Projects/C_elegans/wormpep">http://www.sanger.ac.uk/Projects/C_elegans/wormpep</a> . This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre (chromosome 20 Mapping Group). Further information can be found at: <a href="http://www.sanger.ac.uk/HGP/Chr20">http://www.sanger.ac.uk/HGP/Chr20</a> . RP4-694B14 is from the library RP41-4 constructed by the group of Pieter de Jong. For further details see <a href="http://www.chori.org/hacpac/home.htm">http://www.chori.org/hacpac/home.htm</a> . VECTR: PYPAC2				
FEATURES	This sequence is the entire insert of clone RP4-694B14. The true left end of clone RP41-40-NB is at 116161 in this sequence. The true right end of clone RP41-9616 is at 20586 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., paired quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one W13 subclone; and the assembly was confirmed by restriction digest.				
Source	Location/Qualifiers				
	1..12867				
	/organism="Homo sapiens"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:9606"				
	/chromosome="20"				
	/map="p11.1-11.22"				
	/clone="RP4-694B14"				
	/clone_lib="RP41-4"				
	1..74				
repeat_region	/note="AluJb repeat: matches 194..266 of consensus"				
repeat_region	787..1088				
repeat_region	/note="AluSp repeat: matches 1..303 of consensus"				
repeat_region	1513..1786				
repeat_region	/note="AluSc repeat: matches 1..295 of consensus"				
repeat_region	1818..2372				
repeat_region	/note="11MC3 repeat: matches 7230..7739 of consensus"				
repeat_region	2171..2669				
repeat_region	/note="AluJb repeat: matches 1..298 of consensus"				
repeat_region	2670..3427				









GenCore version 5.1.6  
Copyright (c) 2002-2003 CompuGen Ltd.

ON PROBABILISTIC NUCLEIC SEARCH, USING FIRST-FIT AND BEST-FIT MODELS

Run on: October 20, 2007, 11:54:13 ; Search time 2163 seconds

without a program.  
102.63 million cells, 100,000 cells

Title: US-289 133

perfect 9:8

Sequence : MAA9\_GGMESGCPDGT\_CAP.....SLAYDEASANTFAP1.VMT 178

SECRET

XGAPex	10.0	XGAPex	0.5
YGAPex	10.0	YGAPex	0.5
FGAPex	6.0	FGAPex	7.0
DEGex	6.0	DEGex	7.0

204747 288 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum CP	seq	length	CP	seq	length
2000	0.0000	0.0000	2000	0.0000	0.0000

Cost: \$1000.00  
Minimum Match: 0%  
Maximum Match: 100%  
Losing first 45 minutes

STANDARD 111

```

MODEL = name, ip2n, model, -DEV x15
C:/cmr-115ft51-spool/PSU1/22913/1/mmr 26102200 17492 -k63app query.fastal.1327
OS=GENE-1 QVAT=fastap -SUFFIX=app MINMATCH=11 -MAXCUT=10 -HCOEXT=2
-NC=100000 -START=1 -END=1 -MATRIX=1000000 -TRANS=1000000 -G1 -LIST=45
-LOCAL=100 -THE SCORE=pcr -THE MAX=100 -THE MIN=0 -ALIGN=15 -M-COE=LOCAL
-CUT=100 -NORM=EXL -HAPS=1500 -MINLEN 0 -MAXLEN 20000000
-USER=10000000 -ACGN 1 13745 -MVAL 2302000 17419 6654 -NCNU=6 -ICPU=3
-NC=MMAL -LARGEST=1 -NEG SCORE=0 -WAT -DSIBLOC=100 -LARGLOC
-DEV TIME=120 -WARN TIME=100 -THRE=15 -XMAP=10 -XGAPEXT=10 -ESAF=6
-EPA=10 -YGA=10 -YGAEXT=0.5 -DEL=6 -DELEXT=7

```

THE

Category	GenEmb1
1	gb_ba1
2	gb_hig1
3	gb_in1
4	gb_om1
5	gb_ov1
6	gb_pat1
7	gb_ph1
8	gb_pa1
9	gb_pr1
10	gb_ro1
11	gb_sts1
12	gb_sy1
13	gb_un1
14	gb_vir1
15	em_ba1
16	em_fur1
17	em_hum1
18	em_ina1
19	em_mur1
20	em_om1
21	em_or1
22	em_ov1
23	em_pat1
24	em_ph1
25	em_pl1
26	em_rc1
27	em_sts1
28	em_un1

```

29: em_vit.
30: em_hqo_hum.
31: em_hqo_inv.
32: em_hqo_other.
33: em_hqo_rus.
34: em_hqo_pla.
35: em_hqo_rodi.
36: em_hqo_mani.
37: em_hqo_vit.
38: em_sy.
39: em_hqo_hum.
40: em_hqo_rus.
41: em_hqo_other.

```

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Summary

Result No.	Query	Score	Match	Length	DB	ID	Description
1	918	100.0	822	6	AB015336.1	Human pro	AB015336.1 Sequence
2	918	100.0	862	6	AB015317	Sequence	AB015317 Sequence
3	918	100.0	1071	6	AB017517	Sequence	AB017517 Sequence
4	918	100.0	1071	6	BC0145207	Human nuc	BC0145207 Human nuc
5	918	100.0	1093	6	BC0119136	Human sap1	BC0119136 Human sap1
6	918	100.0	1088	6	AF0116050	Sequence	AF0116050 Sequence
7	918	100.0	1363	6	AF0368829	Homo sap1	AF0368829 Homo sap1
8	918	100.0	1361	6	AF0321617	Homo sap1	AF0321617 Homo sap1
9	913	99.5	1069	9	AF0317629	Homo sap1	AF0317629 Homo sap1
10	903	98.6	789	9	AF0317630	Homo sap1	AF0317630 Homo sap1
11	892	97.2	905	9	AF0317678	Homo sap1	AF0317678 Homo sap1
12	870	94.8	904	10	BC0249478	Homo sap1	BC0249478 Homo sap1
13	870	94.8	904	10	BC0249478	Homo sap1	BC0249478 Homo sap1
14	765	83.3	467	6	BC0108605	EST and e	BC0108605 EST and e
15	615.5	67.2	85234	6	BC0235539	Sequence	BC0235539 Sequence
16	548	60.1	163456	9	AF0110711	Human DNA	AF0110711 Human DNA
17	590	64.3	127867	9	U05694824	Human sap1	U05694824 Human sap1
18	566.5	61.7	27188	2	AC0050663	Homo sap1	AC0050663 Homo sap1
19	566.5	61.7	271364	9	AL0359844	Human DNA	AL0359844 Human DNA
20	519.5	56.6	119561	9	AL0367929	Human DNA	AL0367929 Human DNA
21	492	41.6	128444	9	AP0000725	Homo sap1	AP0000725 Homo sap1
22	397	36.7	173810	9	AF0061602	Homo sap1	AF0061602 Homo sap1
23	319	34.7	233422	2	AC0093358	Homo sap1	AC0093358 Homo sap1
24	319	34.7	249937	2	AC0121198	Rattus no	AC0121198 Rattus no
25	315	34.3	217692	2	AF0092711	Mus muscu	AF0092711 Mus muscu
26	332	31.6	164076	9	AF0210359	Homo sap1	AF0210359 Homo sap1
27	292	30.8	165159	3	AF0069577	Homo sap1	AF0069577 Homo sap1
28	284	30.9	155539	3	AF0069577	Homo sap1	AF0069577 Homo sap1
29	248.5	27.1	753	3	AF0118876	Drosophila	AF0118876 Drosophila
30	248.5	27.1	44815	2	AF0118876	Drosophila	AF0118876 Drosophila
31	248.5	27.1	165704	3	AF0382884	Drosophila	AF0382884 Drosophila
32	248.5	27.1	189285	3	AF0033752	Drosophila	AF0033752 Drosophila
33	196	21.4	205925	3	AC0009711	Homo sap1	AC0009711 Homo sap1
34	134	14.6	164928	9	AF0087752	Homo sap1	AF0087752 Homo sap1
35	134	14.6	165129	9	AF0087752	Homo sap1	AF0087752 Homo sap1
36	123	13.4	232944	2	AC0120382	Mus muscu	AC0120382 Mus muscu
37	119	13.0	502	6	BC0050337	Arabidops	BC0050337 Arabidops
38	119	13.0	603	9	AF0117706	Arabidops	AF0117706 Arabidops
39	113.5	12.4	366	3	AF0211046	Tricholium	AF0211046 Tricholium
40	112.5	12.3	365	3	AF0211049	Human ber	AF0211049 Human ber
41	107.5	11.7	5191	14	AB075774	Human ber	AB075774 Human ber
42	107.5	11.7	5191	14	AB075776	Human ber	AB075776 Human ber
43	107.5	11.7	5626	14	AB075778	Oryza sat	AB075778 Oryza sat
44	107.5	11.7	145905	2	AC132490	Oryza sat	AC132490 Oryza sat
45	107.5	11.7	154967	2	AC135430	Oryza sat	AC135430 Oryza sat

## ALLEGMENTS

511

```

BD015361
LOCUS      BD015361             822 bp     DNA     linear     PAT 27-AUG-2002
DEFINITION Human protein and cDNA[4].
ACCESSION BD015361
VERSION   BD015361.1 GI:22556499
KEYWORDS  JP 2001218584-A/7.
SOURCE    Homo sapiens (human)
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Euteria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE Kato,S. and Saeki,M.
AUTHORS   Kato,S. and Saeki,M.
TITLE     Human protein and cDNA[4].
JOURNAL   JAPAN SCIENCE AND TECHNOLOGY CORP
COMMENT   QS Homo sapiens (human)
          PN JP 2001218584-A/7
          FD 14 AUG-2001
          EF 08-FEB-2000 JP 2000031062
          PI SEIJI KATO,MIHORO SAEKI
          PC C12N15/09,C07K14/475,C07K16/18,C12N1/19,C12N1/19,01R02.01,PC
          CC C12P2/08,C12N15/00,C12N5/30
          FH Key Location/Qualifiers
          FT CDS Location/Qualifiers
              1..822
              /organism="Homo sapiens"
              /mol_type="genomic DNA"
              /db_xref="taxon:9606"
              /35_xref="taxon:9606"
BASE COUNT 210 a 186 c 199 g 227 t
ORIGIN
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```

```

11. 430 CAGGTGCTGGAGGATATCATATGTCAGGACACAAAAGCCCGGACATCCCTCAGAGGCTCT 489
    161 LeuAlaTyrLeuGlnGlnAlaSerAlaAsnIleProAlaPheLeuLysProThr 178
    480 TTGGGCTAGCTGGAAGAGGAGGATGTCGACATCTCTGACATCTCTGACATCTCTGACATCT 543

RESULT 4
LOCUS      AX017517
DEFINITION Sequence 65 from Patent WO/94/04543.
ACCESSION  AX017517
VERSION     AX017517.1  GI:11642714
KEYWORDS   Homo sapiens (human)
SOURCE      Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE   1. (bases 1 to 1271)
AUTHORS     Schwitt,A., Specht,T., Cuhl,E., Wintzman,B., Shomitt,A., Pirarski,C., Cuhl,E. and Rosenthal,A.
TITLE       Human nucleic acid sequences from normal breast tissue
JOURNAL     Patent: WO 99/4765-A 65 23-SEP-1999;
            SCHWITT ARWIN (DE); SPECHT THOMAS (DE); CUHL EDGAR (DE); WINTZMAN BENNO (DE); ROSENTHAL ANDRE (DE); METZGER JES FUES GEN-MERESCHON (DE); HILARSKY CHRISTIAN (DE)
FEATURES    Source: 1071
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="GenBank:W"
            /db_xref="taxon:"9606"
            281 A 218 C 265 G 117 T
            Alignment Statistics:
            Pred. No. 1,246-64 Length: 1271
            Score: 918.00 Matches: 178
            Percent Similarity: 100.0% Conservative: 0
            Best Local Similarity: 100.0% Mismatches: 0
            Query Match: 100.0% Indels: 0
            DB: 6 Gaps: 0
US-10-029-137-2 (1-178) x AX017517 (1-1271)
CY 1 MetAlaAlaProGlyGlyGlyMetPheSerGlyGlnProProGlyGlyProGlnAlaPro 20
Db 10 ATGGCTAGCTCCACACTAGGGGGGTATGTTTTCTGGGAGGACCCCGGTCCTCTCTCAGGCTGG 69
CY 21 ProGlyLeuProGlyGlnAlaSerLeuLeuGlnAlaAlaProGlyValAlaPheArgProSer 40
Db 70 CCGGGGCTTCGGGGCCCAAGCTTCGGCTTCTCAGGACAGGTCGAGGGGCTCTCCAGACCTTCC 129
CY 41 SerSerThrLeuValAspGlnLeuGluSerSerProGlnAlaCysPheAlaSerLeuVal 60
Db 130 AGCAGGACTTCTGGTGACGAGATTCGAGATCATCTTTGAGGCTTCTGCTTCTGCTGCTG 189
CY 61 SerGlnAspTyrValAsnGlyThrAspGlnGlnIleAlaThrGlnGlyValAsnLeuCys 80
Db 130 AATCAGGACTATGTCAATGTCAGCATATCATATGAGAAATTCATACAGCTGCTGATAGAGT 249
CY 81 LeuGlnThrLeuValAspGlnLeuValAlaGlnThrGlnCysProPheLeuLeuLysThrGln 100
Db 130 ATCCAGGAAGTTCTGGATATTGACACACACAGAGAAATGTTTCTTTACAAAAGAATG 309
CY 101 GlnLeuSerValGlnLysProGlnIleValIleLeuGlnAsnValSerGlnLeuArgAsn 120
Db 130 CAGTATATGTCGAGAAACCAAGCAAGATATCAAGAGAGATGTGTTCAGACTATAGGAT 369
CY 121 GluLeuGlnArgLysPheAlaLeuValGlnCysHisIleValIleLysLeuArgHisIleProGln 140
Db 130 GAATTACAGCGGAAGATGTACTATTCTCCAAAGCAATTCACAAAGATTCAGGATTCGGAG 429
CY 141 GlnValLeuGlnAspPheAsnValGlnHisLysLysLysPheAlaAspIleProGlnLysSer 160

```

```

Db 430 CAGGTGCTGGAGGATATCATATGTCAGGACACAAAAGCCCGGACATCCCTCAGAGGCTCT 489
CY 161 LeuAlaTyrLeuGlnGlnAlaSerAlaAsnIleProAlaPheLeuLysProThr 178
    179 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 238
    480 TTGGGCTAGCTGGAAGAGGAGGATGTCGACATCTCTGACATCTCTGACATCTCTGACATCT 543

RESULT 4
LOCUS      RD135207
DEFINITION Human nucleic acid sequence originating in normal mammary tissue.
ACCESSION  RD135207
VERSION     RD135207.1  GI:21210154
KEYWORDS   Homo sapiens (human)
SOURCE      Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE   1. (bases 1 to 1271)
AUTHORS     Specht,T., Wintzman,B., Shomitt,A., Pirarski,C., Cuhl,E. and Rosenthal,A.
TITLE       Human nucleic acid sequence originating in normal mammary tissue
JOURNAL     Patent: JP 2002506639-A 54 05-MAR-2002;
            METZGER GEBELUSCHAF FUER GENOME FORSCHUNG MSH
            CS Homo sapiens (human)
            IN JP 2002506639-A/54
            PD 05-MAR-2002
            PF 19 MAR-1999 JP 2002516438
            PR 20-MAR-1996 DE 194 11 935 0
            P1 THOMAS SPECHT;BENNO WINTZMAN;ARWIN SHOMITT;CHRISTIAN PIRARSKI;
            P2 EDGAR CUHL;
            P3 ANDRE ROSENTHAL
            PC C22N15/09,A61K45/00,A61P35/00,A61P43/00,A61P43/50,C02K14/47,
            PC C07K16/15,
            PC C12N9/10,C12P21/02,C12P21/02,C12P21/08,G01N33/65//A61K18/00, PC
            C12N16/00,
            PC C22N15/00,A61K23/00
            CC Human nucleic acid sequence originating in normal mammary CC
            TT tissue
            FH KEY Location/Qualifiers
            FT source 1..1271
            FT /organism="Homo sapiens (human)"
            FT /molecule="genomic DNA"
            FT /db_xref="taxon:"9606"
            BASE COUNT 241 A 218 C 265 G 117 T
            ORIGIN
            1
            1271
            Location/Qualifiers
            1..1271
            /organism="Homo sapiens"
            /molecule="genomic DNA"
            /db_xref="taxon:"9606"
            Alignment Statistics:
            Pred. No. 1,246-64 Length: 1271
            Score: 918.00 Matches: 178
            Percent Similarity: 100.0% Conservative: 0
            Best Local Similarity: 100.0% Mismatches: 0
            Query Match: 100.0% Indels: 0
            DB: 6 Gaps: 0
US-10-029-137-2 (1-178) x RD135207 (1-1271)
CY 1 MetAlaAlaProGlyGlyGlyMetPheSerGlyGlnProProGlyGlyProGlnAlaPro 20
Db 10 ATGGCTAGCTCCACACTAGGGGGGTATGTTTTCTGGGAGGACCCCGGTCCTCTCTCAGGCTGG 69
CY 21 ProGlyLeuProGlyGlnAlaSerLeuLeuGlnAlaAlaProGlyValAlaPheArgProSer 40
    41 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 48
    70 CCGGGGCTTCGGGGCCCAAGCTTCGGCTTCTCAGGACAGGTCGAGGGGCTCTCCAGACCTTCC 129
CY 41 SerSerThrLeuValAspGlnLeuGluSerSerProGlnAlaCysPheAlaSerLeuVal 60
    61 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 68
    130 AGCAGGACTTCTGGTGACGAGATTCGAGATCATCTTTGAGGCTTCTGCTTCTGCTGCTG 189
CY 61 SerGlnAspTyrValAsnGlyThrAspGlnGlnIleAlaThrGlnGlyValAsnLeuCys 80
    81 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 88
    130 AATCAGGACTATGTCAATGTCAGCATATCATATGAGAAATTCATACAGCTGCTGATAGAGT 249
CY 81 LeuGlnThrLeuValAspGlnLeuValAlaGlnThrGlnCysProPheLeuLeuLysThrGln 100
    101 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 108
    130 ATCCAGGAAGTTCTGGATATTGACACACACAGAGAAATGTTTCTTTACAAAAGAATG 309
CY 101 GlnLeuSerValGlnLysProGlnIleValIleLeuGlnAsnValSerGlnLeuArgAsn 120
    121 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 128
    130 CAGTATATGTCGAGAAACCAAGCAAGATATCAAGAGAGATGTGTTCAGACTATAGGAT 369
CY 121 GluLeuGlnArgLysPheAlaLeuValGlnCysHisIleValIleLysLeuArgHisIleProGln 140
    141 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 148
    130 GAATTACAGCGGAAGATGTACTATTCTCCAAAGCAATTCACAAAGATTCAGGATTCGGAG 429
CY 141 GlnValLeuGlnAspPheAsnValGlnHisLysLysLysPheAlaAspIleProGlnLysSer 160

```















BASE COUNT 116 a 113 c 122 g 108 t 6 others

# Alignment Scores:

red. No.: 5,29e+54 length: 465  
 core: 742/100 matches: 344  
 percent similarity: 97.30% consorvative: 3  
 est local similarity: 97.30% mismatches: 4  
 query match: 80.61% indels: 3  
 gaps: 6

S-10-029 137 2 1 178) x 8025539 11-465:

1 MetAlaAlaProLeuGlyGlyMetPreSerGlyValProProGlyProProGlnAlaPro 20  
 20 ATGCGCGCTCCATAGCGGATGTTTCTGCGACACCGACCGTCCCTCTCAGGCGCCG 79  
 21 ProGlyLeuProGlyGlnAlaSerLeuLeuGlnAlaAlaProGlyAlaProGlyProSer 40  
 22 CTGACATCTTCCGCGCCGACGCTTCTCTCAGACAGCTTCTAGATTTCTTACATTTCT 139  
 41 SerSerThrLeuValAspGluLeuGluSerSerLeuGlnAlaGlySerLeuVal 60  
 140 ACCATTAATCTTCCGACGATCGGATATCTGAGATTCAGATTCAGATTCAGATTCAG 149  
 61 SerGluAspTyrValAsnGlyThrAspGlnGluGluLeuThrValAspGlnCys 80  
 260 ATTCAGAGATGATGTCAGATGACGATGACGATGACGATGACGATGACGATGACGAT 259  
 81 IleCysPheSerLeuAspGluLeuAspGlnGluSerSerLeuValAspGlnCys 100  
 260 ATTCAGAGATGATGTCAGATGACGATGACGATGACGATGACGATGACGATGACGAT 319  
 101 GlnLeuSerValGlnPheProGluGlnValLeuGlyGluAspValSerGluLeuAsn 120  
 320 CAATTCATCTCTCAGAACACAGACAGCAAGTATTAACATGATCTGTCAGAACTAGGA 319  
 121 GlnLeuGluArgLysAspAlaLeuValGlnPheSerLeuThrValAspGlnHisTyrGln 140  
 380 GAATTCAGCGCGGACGATGACGATGACGATGACGATGACGATGACGATGACGATGAC 439  
 141 GlnValLeuGlnAspGlnAsnVal 148  
 440 CAGTGTCTGACGACATCAAGCTG 463

HS22725 55304 bp DNA linear PRI 22 MAY 2002  
 Human DNA sequence from clone B56-22025 on chromosome Xp11.22-11.3,  
 complete sequence.

ACCESSION AC031849

VERSION 2 GI:21212870

KEYWORDS RTG;

SOURCE Homo sapiens

ORGANISM Homo sapiens (human)

REMARKS: Eukaryote; Vertebrate; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Chordata; Mammalia; Homo;

1 clones 1 to 85304;

Direct Submission

Submitted (21-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,

Cambridgehire, CB0 0SA, UK. E-mail: enquiries;

humquery@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk

On May 25, 2002 this sequence version replaced gi13950350.

During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >=

20) an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one phasmid subclone of more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. The following  
 abbreviations are used as associate primary accession numbers given  
 in the feature table with their source databases: EMBL, EMBL, SRA,  
 SRA, SRA, SRA, SRA, SRA, SRA, SRA, SRA, SRA, SRA, SRA, SRA, SRA,  
 database can be found at:

http://www.sanger.ac.uk/Projects/C\_Elegans/wormseq/ This sequence

was generated from part of bacteriophage contigs of human

chromosome X, constructed by the Sanger Centre Chromosome X Mapping

Group. Further information can be found at:

http://www.sanger.ac.uk/EMBL/EMBL

EMBL database is from the library EMBL 4, constructed by the group of

Philip de Jong. For further details see

http://www.sanger.ac.uk/EMBL/EMBL

EMBL 4, EMBL 4.

## FEATURES

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

Source

DB 5939 AGGGCTCTGGCCCTAGCTTAGAGAGGCGATCTGGCGATATTCCTCCACCTGTGAGGCCAA 5939  
24 128 hr 17s  
DB 5939 00 5937

Search completed: October 21, 2003, 11:34:19  
Run time: 1.38s secs







```

CC 21 ProGlyLeuProGlyGlnAlaSerLeuGlnAlaAlaSerLeuGlnAlaSerLeu 40
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC 22 CCGGGGTTCCGGGAAAGGTTTATTCAGAGAGAGAGAGAGAGAGAGAGAGAG 129
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC 41 SerSerThrLeuValAspLeuGlnAlaSerLeuGlnAlaSerLeuGlnAla 60
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC 42 ATAGAGTACTGGTTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAG 149
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC 43 SerSerThrLeuValAspLeuGlnAlaSerLeuGlnAlaSerLeuGlnAla 80
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC 44 ASPAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 239
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC 45 ThrGlnSerThrLeuValAspLeuGlnAlaSerLeuGlnAlaSerLeuGln 120
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC 46 ATAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTC 309
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC 47 GlnGlnSerValGlnGlnSerLeuGlnAlaSerLeuGlnAlaSerLeuGln 120
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC 48 CAGTATATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTC 389
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC 49 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 140
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC 50 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 429
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC 51 GlnValLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 120
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC 52 CAGTATATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTCAGAGATTC 489
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC 53 LeuLeuLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 179
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
CC 54 ThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 549
CC ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

AA 01 14 standard: cDNA; 465 BP.

AA 01 14:

AA 01 14: (first entry)

Human secreted protein 5' EST, 582 ID M11799

Human 5' EST, expressed sequence from secreted protein; cDNA isolation; gene therapy; chromosome mapping; ss.

Human sapiens.

EP:033401-A2.

06-SEP-2003.

21 FEB 2000; 2000EP-020610.

26 FEB 1999; 199US-0122429.

GENSET.

Dr. Mas Milne Edwards C, Durlent A, Guindard J.

WPI:0000 500381/45.

P-18086; AAGC1788.

New nucleic acid that is a 5' expressed sequence tag (5' EST) for obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for diagnostic, forensic, gene therapy and chromosome mapping procedures -

Claim 1: SEQ ID 1792; 71pp + CD ROM; English.

The present sequence is one of a large number of 5' ESTs derived from mammal encoding secreted proteins. An EST has been identified within the genome. The 5' ESTs were prepared from total human RNAs or polyA-RNAs derived from 10 different tissues. EST sequences usually correspond

mainly to the 3' untranslated region (UTR) of the mRNA because they are often obtained from oligo-dT primed cDNA libraries. Such ESTs are not well suited for isolating cDNA sequences derived from the 5' ends of mRNAs and even in those cases where longer cDNA sequences have been obtained, the full 5' UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and chromosome mapping procedures. They are used to obtain upstream regulatory sequences and to design expression and secretion vectors.

Sequence 465 BP: 116 A; 115 C; 122 G; 108 T; 6 other:

Alignment Scores:

Align. No.	Score	Length	Matches
1	940.00	465	144
2	97.30%	Conservative:	3
3	97.30%	Mismatches:	4
4	90.61%	Indels:	3
5	21	Gaps:	3

US 10 029 137-2 (1-179) x AAGC1794 (1-465)

QY 1 MetAlaAlaProLeuGlyGlyMetPheSerGlyGlnProProGlyProPheGlnAlaAla 20

DB 20 ATGGGGGTCCTACTAGGCGGATGTTCTGGGACGCCCGGTCCTCCCTCAGGCGCG 29

QY 21 ProGlyLeuProGlyGlnAlaSerLeuGlnAlaAlaProGlyValaProAspProSer 40

DB 80 CCGGGGCTCCCGGCGACCTTCCTCTTCAGGTCAGCTCCAGGCGCTCTCAGACCTTC 139

QY 41 SerSerThrLeuValAspLeuGlnGlnSerPheGlnAlaCysPheAlaSerLeuVal 60

DB 140 AGCAATATCTTGGGACGAGTTGCAATCATCTTTCAGGCTTCTTSCATCTCTGTG 199

QY 61 SerGlnAspGlyValAspGlyThrAspGlnGlnGlnLeuAspThrGlyValAspGln 80

DB 200 ATTCGAGCTATGTCATGCGACCGTACCGATCAGAGAGAAATCGAACCGGTCTTCAGT 259

QY 81 IleGlnLysThrLeuAspGlnAlaArgGlnThrGlnCysPhePheLeuGlnLysArg 100

DB 260 ATCCAGAAAGNCTCAGATATTCAGACAGACAGACAGATGTTTCTTTCACAAAAGATG 319

QY 101 GlnGlnSerValGlnGlnSerLeuGlnValIleLysGlnAspValSerGlnLeuArg 120

DB 300 TACTTTCCTCCAGAGAGAGAGAGATTCACAAAGAGGATGTCAGAGAAATAGAGAAAT 379

QY 121 GlnLeuGlnAlaGlnAspAlaLeuValGlnLysHisLeuThrLysLeuAlaHisSerP 140

DB 390 GAATATACCGGAGAGATTCAGTCTCCAGAGAGCTTCACAAAGCTCAGGATTCGGTAA 419

QY 141 GlnValLeuGlnAspGlnLeuVal 148

DB 440 CAGTCTCCAGGACATCAGCTG 463

RESULT 4

AAS76383

10 AAS76383 standard: cDNA; 1941 BP.

XX AC AAS76383;

XX DT 13-FEB-2002 (first entry)

XX DE cDNA encoding novel human diagnostic protein #12187.

XX DE Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX CS Homo sapiens.

XX CS NC200175067-A2.

XX XX 11-OCT-2001.





The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is one of the probes of the invention.

Sequence 216 BP: 39 A: 55 C: 60 G: 62 T: 3 Other:

XX  
PS  
ES  
CC  
CC  
CC  
CC  
CC  
CC  
CC

Example 4: SEQ ID NO. 1164; 859pp - Sequence listing; English.

The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human bone marrow. They can be used to measure gene expression in bone marrow samples, which may enable the improved diagnosis and treatment of cancers such as lymphoma, leukemia and myeloma. The present sequence is one of the probes of the invention.

XX  
PS Example 4: SEQ ID NO. 1164; 859pp - Sequence listing; English.  
CC  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukemia and myeloma. The present sequence is one of  
CC the probes of the invention.



XX (MOLECULAR DYNAMICS INC.)  
 XX Penn SG, Hanzel DK, Chen W, Rank PR;  
 XX WPI; 2001-476286/51.  
 XX Novel single exon nucleic acid probe used to measuring gene expression  
 XX in a human breast  
 XX Claim 25: SEQ ID NO 5123; 322bp; English.  
 XX The present invention relates to novel single exon nucleic acid probes.  
 XX The present sequence is one such probe. The probes are useful for  
 XX measuring human gene expression in a human breast sample, where the probe  
 XX hybridizes at high stringency to a nucleic acid expressed in the human  
 XX breast. The probes are useful for predicting, diagnosing, grading,  
 XX staging, monitoring and prognosing diseases of the human breast,  
 XX particularly those diseases with polygenic aetiology. The diseases  
 XX include breast cancer, disorders of development, inflammatory diseases and  
 XX non-carcinoma tumours.  
 XX Note: The sequence data for this patent did not form part of the printed  
 XX specification, but was obtained in electronic format directly from WPI  
 XX at ftp.wpi.edu/pub/published\_pat\_sequences.  
 XX Sequence 216 BP; 13 A; 55 C; 60 G; 62 T; 0 other;  
 XX  
 XX Alignment Scores:  
 XX Predicted No. 7.46 23 Length: 216  
 XX Score: 357.00 Matches: 65  
 XX Percent Similarity: 100.00% Conserved: 1  
 XX Best Local Similarity: 98.48% Mismatch: 2  
 XX Query Match: 96.71% Indels: 0  
 XX Gaps: 0  
 XX  
 XX US-10 029-137-2 11 1741 x AA105132 11-2161  
 XX  
 XX 113 GluAspValSerGluLeuArgAspGluLeuGluArgGlyAspAlaLeuValGlnIleHis 130  
 XX 130 CAGGAGGTGTGAGAACCTAAGGATGATACAGGGAAGATGTCATAGTCCAGAGCAC 142  
 XX 143 LeuThrLysLeuArgHisTrpGlnGluValLeuGluAspLeuAsnValSerHisLysLys 162  
 XX 161 TGGAGAGCTGAGACCATTCACACAGCTCTCTCTGAGCATTAACATTTACACGACAAAG 82  
 XX 81 CCGTGGACATCCCTCAGGCTCTCTTGGTACTGTGGATAGACATTTCCACATCTCT 22  
 XX 173 AlaProLeuLysProThr 178  
 XX 21 GCACTCTGTAGAGCAACG 4  
 XX  
 XX RESULT 12  
 XX ABS12673/C  
 XX ID AR38171 standard; DNA; 216 BP.  
 XX AC AR38171  
 XX CC AR38171  
 XX 25 FEB 2003 (first entry)  
 XX Human liver single exon probe, SEQ ID NO 13161.  
 XX  
 XX Human; single exon nucleic acid probe; liver; cirrhosis;  
 XX hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;  
 XX coronary heart disease; ss.  
 XX  
 XX Homo sapiens.  
 XX  
 XX WC200157273-A2.  
 XX 09-AUG-2001.  
 XX

XX 30-JAN-2001; 2001US-0017444.  
 XX  
 XX 04-FEB-2001; 2002US-0189117  
 XX 26-MAY-2001; 2002US-0237456  
 XX 03-JUN-2001; 2002US-0698404  
 XX 03-AUG-2001; 2002US-0693166  
 XX 21-SEP-2001; 2002US-0834657  
 XX 27-SEP-2001; 2002US-0736159  
 XX 04-OCT-2001; 2002US-0034261.  
 XX (MOLECULAR DYNAMICS INC.)  
 XX Penn SG, Hanzel DK, Chen W, Rank PR;  
 XX WPI; 2001-476286/51.  
 XX Human genome derived single exon nucleic acid probes useful for  
 XX analysing gene expression in human liver tissue.  
 XX Claim 4: SEQ ID NO 13161; 322bp; English.  
 XX The invention relates to a method of measuring a gene product in a  
 XX measuring human gene expression in a sample derived from a human liver  
 XX liver, comprising one or more defined nucleic acid sequences given in the  
 XX specification or equivalents/fragments of the probe hybridises at high  
 XX stringency to a nucleic acid molecule expressed in the human adult  
 XX liver. It may be used for predicting, measuring and displaying gene  
 XX expression in samples derived from human adult liver. The genes  
 XX identified may be involved in genetic liver diseases such as cirrhosis,  
 XX hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which  
 XX is associated with coronary heart disease. ABS12673: ABS12673 represents  
 XX human liver single exon nucleic acid probes of the invention.  
 XX Note: The sequence information for this patent does not appear in the  
 XX printed specification but was obtained in electronic format directly  
 XX from WPI at ftp.wpi.edu/pub/published\_pat\_sequences.  
 XX  
 XX Sequence 216 BP; 13 A; 55 C; 60 G; 62 T; 0 other;  
 XX  
 XX Alignment Scores:  
 XX Predicted No. 7.46 23 Length: 216  
 XX Score: 357.00 Matches: 65  
 XX Percent Similarity: 100.00% Conserved: 1  
 XX Best Local Similarity: 98.48% Mismatch: 2  
 XX Query Match: 96.71% Indels: 0  
 XX Gaps: 0  
 XX  
 XX US-10 029 137 2 11 1741 x ABS12673 11-2161  
 XX  
 XX 113 GluAspValSerGluLeuArgAspGluLeuGluArgGlyAspAlaLeuValGlnIleHis 130  
 XX 130 CAGGAGGTGTGAGAACCTAAGGATGATACAGGGAAGATGTCATAGTCCAGAGCAC 142  
 XX 143 LeuThrLysLeuArgHisTrpGlnGluValLeuGluAspLeuAsnValSerHisLysLys 162  
 XX 161 TGGAGAGCTGAGACCATTCACACAGCTCTCTCTGAGCATTAACATTTACACGACAAAG 82  
 XX 81 CCGTGGACATCCCTCAGGCTCTCTTGGTACTGTGGATAGACATTTCCACATCTCT 22  
 XX 173 AlaProLeuLysProThr 178  
 XX 21 GCACTCTGTAGAGCAACG 4  
 XX  
 XX RESULT 13  
 XX ABS12673/C  
 XX ID ABS12673 standard; DNA; 216 BP.  
 XX AC ABS12673  
 XX CC ABS12673  
 XX 19-AUG-2002 (first entry)  
 XX



X The invention relates to a single exon nucleic acid probe for  
 C measuring human gene expression in a sample derived from human foetal  
 C liver. The single exon nucleic acid probes may be used for predicting,  
 C measuring and displaying gene expression in samples derived from human  
 C fetal liver. The present sequence is a single exon nucleic acid  
 C probe of the invention.  
 C Note: The sequence data for this patent did not form part of the  
 C printed specification, but was obtained in electronic format directly  
 C from WIPO at ftp.wipo.int/pub/published-pat-sequences.  
 X Sequence 478 BP: 136 A; 111 C; 100 G; 131 T; 0 other;

Alignment Scores:  
 2,33e-22 Length: 478  
 137.00 Matches: 65  
 100.00% Conservative: 1  
 94.48% Mismatches: 0  
 36.71% Indels: 0  
 22 Gaps: 0

US-10-029 137 2 1 136 A; 111 C; 100 G; 131 T; 0 other;  
 133 GluAspValSerLeuGluLeuValGlnGlyAspAlaLeuValGlnLysHis 132  
 470 CAGATCTCTTCACACTAGCAATGATATTACATGCGGAGGATGCGATGCGACGCA 411  
 133 LeuThrLysLeuArgHisTrrpGlnGlnValLeuGluAspIleAsnValHisCysLys 152  
 410 CTGACAAGCTGAGCGATTGACGACGATGCTGAGGACATCAACCTGACGACAAAG 411  
 153 ProLysAspIleProGlnGlySerLeuAlaTyrLeuGlnAlaSerAlaAsnIlePro 172  
 430 CCGCGCGGAATCCCTGAGGCTCTTTCTTACTCTGACGAGGATCTGGTAACATCT 291  
 133 AlaProGlnLysProThr 178  
 430 CAGATCTCTTCACACTAGCAATGATATTACATGCGGAGGATGCGATGCGACGCA 411  
 133 AlaProGlnLysProThr 178  
 430 CAGATCTCTTCACACTAGCAATGATATTACATGCGGAGGATGCGATGCGACGCA 411

RESULT 15  
 133 GluAspValSerLeuGluLeuValGlnGlyAspAlaLeuValGlnLysHis 132  
 470 CAGATCTCTTCACACTAGCAATGATATTACATGCGGAGGATGCGATGCGACGCA 411  
 133 LeuThrLysLeuArgHisTrrpGlnGlnValLeuGluAspIleAsnValHisCysLys 152  
 410 CTGACAAGCTGAGCGATTGACGACGATGCTGAGGACATCAACCTGACGACAAAG 411  
 153 ProLysAspIleProGlnGlySerLeuAlaTyrLeuGlnAlaSerAlaAsnIlePro 172  
 430 CCGCGCGGAATCCCTGAGGCTCTTTCTTACTCTGACGAGGATCTGGTAACATCT 291  
 133 AlaProGlnLysProThr 178  
 430 CAGATCTCTTCACACTAGCAATGATATTACATGCGGAGGATGCGATGCGACGCA 411

AAK00050 standard; DNA: 478 BP.  
 AAK00050:  
 05-NOV 2001 (first entry;  
 Human: brain expressed single exon probe SEQ ID NO: 41.  
 Human: brain expressed exon; gene expression; analysis; probe;  
 microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
 epilepsy; cancer; ss.  
 Home captions:  
 WO200157275-A2.  
 09 AUG 2001.  
 30-JAN 2001; 2001WO-US00667.  
 04 FEB-2003; 2003US-0180322.  
 26-MAY 2003; 2003US-0207456.  
 30-JUN-2003; 2003US-0608408.  
 03-AUG-2003; 2003US-0623666.  
 21-SEP-2003; 2003US-0234687.  
 27-SEP-2003; 2003US-0236359.  
 04-OCT-2003; 2003US-0024263.  
 (MOLECULAR DYNAMICS INC  
 Penn SG, Hanzel DK, Chen W, Rank DR;

DS WP1: 2301-483446/52.  
 XX Single exon nucleic acid probes for analyzing gene expression in human  
 PT brains -  
 XX  
 XX Example 4: SEQ ID NO: 41: 650bp + Sequence Listing: English.  
 XX  
 CC The present invention provides a method of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC brain. They can be used to measure gene expression in brain cell samples,  
 CC which may enable the diagnosis and improved treatment of nervous system  
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
 CC epilepsy and cancers. The present sequence is one of the probes of the  
 CC invention.  
 XX  
 XX Sequence 478 BP: 136 A; 111 C; 100 G; 131 T; 0 other;  
 SQ

Alignment Scores:  
 2,33e-22 Length: 478  
 137.00 Matches: 65  
 100.00% Conservative: 1  
 94.48% Mismatches: 0  
 36.71% Indels: 0  
 22 Gaps: 0

US-10-029 137 2 1 136 A; 111 C; 100 G; 131 T; 0 other;  
 133 GluAspValSerLeuGluLeuValGlnGlyAspAlaLeuValGlnLysHis 132  
 470 CAGATCTCTTCACACTAGCAATGATATTACATGCGGAGGATGCGATGCGACGCA 411  
 133 LeuThrLysLeuArgHisTrrpGlnGlnValLeuGluAspIleAsnValHisCysLys 152  
 410 CTGACAAGCTGAGCGATTGACGACGATGCTGAGGACATCAACCTGACGACAAAG 411  
 153 ProLysAspIleProGlnGlySerLeuAlaTyrLeuGlnAlaSerAlaAsnIlePro 172  
 430 CCGCGCGGAATCCCTGAGGCTCTTTCTTACTCTGACGAGGATCTGGTAACATCT 291  
 133 AlaProGlnLysProThr 178  
 430 CAGATCTCTTCACACTAGCAATGATATTACATGCGGAGGATGCGATGCGACGCA 411

Search completed: Oct 21, 2003, 00:00:44  
 Job time: 25.9 secs



GenCore version 3.1.6  
Copyright (c) 1993-1999 CompuLink Ltd.

M protein protein search, using sw tools.

Run on: October 20, 2003, 13:21:57 / Search time 47 seconds  
(without alignment)  
124.751 Million cell updates/sec

Title: us-10-029-137-2

Sequence: 1 WAAHNDNENFQPPSHAP...USAYACASAKAPKPT 178

Sequencing lab: BDO5062

Gap: 10.0, Gapped: 0

Searches: 1107863 seqs, 14826501 residues

Total number of hits satisfying chosen parameters: 1107863

Maximum hit seq length: 0

Maximum hit seq length: 200600000

Fast p: 0.000000 Minimum Match: 0%

Maximum Match: 100%

Hitting time: 45 seconds

Database: 1 A\_Genescp\_12Jun01\*

- 1 /SIDS:/gcdata/genescp/genescp-emb/AA1981.DAT\*
- 2 /SIDS:/gcdata/genescp/genescp-emb/AA1981.DAT\*
- 3 /SIDS:/gcdata/genescp/genescp-emb/AA1982.DAT\*
- 4 /SIDS:/gcdata/genescp/genescp-emb/AA1982.DAT\*
- 5 /SIDS:/gcdata/genescp/genescp-emb/AA1984.DAT\*
- 6 /SIDS:/gcdata/genescp/genescp-emb/AA1984.DAT\*
- 7 /SIDS:/gcdata/genescp/genescp-emb/AA1985.DAT\*
- 8 /SIDS:/gcdata/genescp/genescp-emb/AA1986.DAT\*
- 9 /SIDS:/gcdata/genescp/genescp-emb/AA1987.DAT\*
- 10 /SIDS:/gcdata/genescp/genescp-emb/AA1988.DAT\*
- 11 /SIDS:/gcdata/genescp/genescp-emb/AA1989.DAT\*
- 12 /SIDS:/gcdata/genescp/genescp-emb/AA1990.DAT\*
- 13 /SIDS:/gcdata/genescp/genescp-emb/AA1991.DAT\*
- 14 /SIDS:/gcdata/genescp/genescp-emb/AA1992.DAT\*
- 15 /SIDS:/gcdata/genescp/genescp-emb/AA1993.DAT\*
- 16 /SIDS:/gcdata/genescp/genescp-emb/AA1994.DAT\*
- 17 /SIDS:/gcdata/genescp/genescp-emb/AA1994.DAT\*
- 18 /SIDS:/gcdata/genescp/genescp-emb/AA1997.DAT\*
- 19 /SIDS:/gcdata/genescp/genescp-emb/AA1998.DAT\*
- 20 /SIDS:/gcdata/genescp/genescp-emb/AA1999.DAT\*
- 21 /SIDS:/gcdata/genescp/genescp-emb/AA2000.DAT\*
- 22 /SIDS:/gcdata/genescp/genescp-emb/AA2001.DAT\*
- 23 /SIDS:/gcdata/genescp/genescp-emb/AA2002.DAT\*
- 24 /SIDS:/gcdata/genescp/genescp-emb/AA2003.DAT\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	918	100.0	181	20	AAV76513 Human ovarian tumor
2	918	100.0	161	20	AAV45526 Human breast tumor
3	740	80.6	148	21	ASG01788 Human secreted pro
4	643.5	70.1	646	22	ASG1136 Novel human diapo
5	335	36.5	65	22	ASG1404 Human liver peptid
6	335	36.5	65	22	ASG2402 Human peptide #53
7	335	36.5	65	22	AB312511 Human peptide #57
8	335	36.5	65	22	AAV93161 Human brain expres
9	335	36.5	65	22	AAV65761 Human bone marrow

10	335	36.5	65	22	AAV1421 Peptide #45 encode
11	335	36.5	65	22	AAV26018 Peptide #55 encode
12	335	36.5	65	22	AAV01771 Peptide #53 encode
13	335	36.5	65	23	ASG3191 Human peptide: esco
14	326.5	35.6	231	22	ASG15102 Novel human diapo
15	249.5	27.1	189	22	ASB51275 Drosophila melanog
16	119	13.0	171	21	AAV21209 Arabidopsis thalia
17	102.5	11.2	422	21	AAV36649 Human prostate can
18	98.5	10.7	1078	22	AAV33100 Human polypeptide
19	98.5	10.7	1089	22	AAV3212 Human protein seq
20	94.5	10.7	1089	22	AAV41386 Human polypeptide
21	93.5	10.7	1089	22	AAV3503 Human human diapo
22	93.5	10.7	1089	21	AAV53127 Human circadian rh
23	96.5	10.5	648	23	ABG6455 Human kinase #6
24	96.5	10.5	648	24	ABG6455 Human kinase #6
25	96.5	10.5	654	23	ABG6454 Human novel kinase
26	96.5	10.5	654	24	ABG6459 Human novel kinase
27	96.5	10.5	741	23	AAV19146 Human kinase #5
28	96.5	10.5	864	23	ABG6451 Human kinase polyp
29	96.5	10.5	864	24	ABG6451 Human novel kinase
30	96.5	10.5	870	23	ABG6450 Human novel kinase
31	96.5	10.5	870	24	ABG6450 Human novel kinase
32	96.5	10.5	889	23	ABG6459 Human novel kinase
33	96.5	10.5	889	24	ABG6459 Human novel kinase
34	96.5	10.5	895	23	ABG6458 Human novel kinase
35	96.5	10.5	895	24	ABG6458 Human novel kinase
36	96.5	10.5	951	23	ABG6467 Human novel kinase
37	96.5	10.5	951	24	ABG6467 Human novel kinase
38	96.5	10.5	957	23	ABG6466 Human novel kinase
39	96.5	10.5	957	24	ABG6466 Human novel kinase
40	96.5	10.5	976	23	ABG6463 Human novel kinase
41	96.5	10.5	976	24	ABG6463 Human novel kinase
42	96.5	10.5	982	23	ABG6462 Human novel kinase
43	96.5	10.5	982	24	ABG6462 Human novel kinase
44	95.5	10.4	205	23	ABD04730 Human -Perj-23 pro
45	95.5	10.4	999	22	ABE6737 Drosophila melanog

ALIGNMENTS

RESULT 1  
AAV76513 standard; Protein: t51 AA.  
AC AAV76513  
CT 10-APR-2000 (first entry)  
XX Human ovarian tumor EST fragment encoded protein 9.  
XX Expressed sequence tag; EST; human; ovarian tumor; anticancer;  
XX gene therapy; treatment.  
XX Homo sapiens.  
XX DEL9817557-A1.  
XX 21-OCT-1999.  
XX 09-APR-1998; 98DE-1017557.  
XX 09-APR-1998; 98DE-1017557.  
XX (META-) METAGEN GES GENOMFORSCHUNG MSH.  
XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;  
XX WPI; 1999-551920/51.  
XX R-PSDB; AA277452.  
XX New nucleic acid sequences expressed in ovarian, and some other, cancer  
XX tissues, and derived polypeptides, for treatment of ovarian cancer and









```

Y      174 PLKPT 176
b      61 PLKPT 65

RESULT 20
D      AAM13621 standard; Protein; 65 AA.
X      AAM13621
X      X
X      T 12-OCT-2001 (first entry)
X      E Peptide #55 encoded by probe for measuring cervical gene expression.
X      W Probe; human; microarray; gene expression; cervical epithelial cells;
W      cervical cancer.
X      X Homo sapiens.
S      X
S      WC200159274 AD
X      X
X      C9-AUG-2001.
D      X
D      30-JAN-2001; 2001NC-0200670.
X      X
X      C4-FEB-2000; 2000US-0180112.
X      R 26-MAY-2000; 2000US-0207456.
X      R 30-JUN-2000; 2000US-0308408.
X      R C3-AUG-2000; 2000US-0632166.
X      R 21-SEP-2000; 2000US-0234687.
X      R 27-SEP-2000; 2000US-0236359.
X      R C4-OCT-2000; 2000SB-0024263.
X      X
X      (MOLE-) MOLECULAR DYNAMICS INC.
A      X
A      Penn SG, Hatzel DK, Chen W, Rank DS;
X      WPI; 2001-489977/53.
X      X
X      Human genome-derived single exon nucleic acid probes useful for
X      analyzing gene expression in human placenta.
X      X
X      Claim 27; SEQ ID No 18447; 487bp; English.
X      X
X      The present invention relates to human single exon nucleic acid probes
X      (SENPs; see AAM13621 standard). The present sequence is a peptide encoded
X      by one such probe. The SENPs are derived from human HeLa cells. The SENPs
X      can be used to produce a single exon microarray, which can be used for
X      measuring human gene expression in a sample derived from human cervical
X      epithelial cells. By measuring gene expression, the probes are therefore
X      useful in grading and/or staging of diseases of the cervix, notably
X      cervical cancer.
X      C Note: The sequence data for this patent did not form part of the printed
X      specification, but was obtained in electronic format directly from WIPO
X      at ftp://patent.us/pat/published_pat_sequences.
X      X
X      Query Match 36.5%, Score 315; DB 22; Length 65;
X      Best Local Similarity 100.0%; Pct. No. 2 66-23;
X      Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
X      X
X      114 DVSELRNEQKQKALVQKHKLKRWKQGVLEDFNVQHKKPAD:POGSLAYEQASANTIPA 174
X      1 DVSELRNEQKQKALVQKHKLKRWKQGVLEDFNVQHKKPAD:POGSLAYEQASANTIPA 60
X      X
X      174 PLKPT 176
X      61 PLKPT 65

RESULT 11
X      X
X      Query Match 36.5%, Score 315; DB 22; Length 65;
X      Best Local Similarity 100.0%; Pct. No. 2 66-23;
X      Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
X      X
X      114 DVSELRNEQKQKALVQKHKLKRWKQGVLEDFNVQHKKPAD:POGSLAYEQASANTIPA 174
X      1 DVSELRNEQKQKALVQKHKLKRWKQGVLEDFNVQHKKPAD:POGSLAYEQASANTIPA 60
X      X
X      174 PLKPT 176
X      61 PLKPT 65

RESULT 12
X      X
X      AAM1371 standard; Protein; 65 AA.
X      X
X      AAM1371
X      X
X      C9-OCT-2001 (first entry)
X      X
X      Peptide #55 encoded by probe for measuring human breast gene expression.
X      DE Probe; human; breast cancer; developmental disorder;
X      X inflammatory disease; proliferative breast disease; non-carcinoma tumor.
X      X
X      X

```

XX Humo sapiens.  
XX W02001:56001 A2.  
XX 15-AUG-2001.  
XX 29 JAN 2001; 2001WO-US00461.  
XX 04 FEB 2000; 2000US-2190112.  
XX 26 MAY 2000; 2000US-2267456.  
XX 16 JUN 2000; 2000US-2629408.  
XX 01 APR 2000; 2000US-2632166.  
XX 21 SEP 2000; 2000US-0214587.  
XX 27 FEB 2000; 2000US-0236359.  
XX 04 OCT 2000; 2000US-0004263.  
XX MOLECULAR DYNAMICS INC.  
XX Peen SG, Hanzel DK, Chen W, Rank DS;  
XX WPI; 2001:476286/51.  
XX Novel single exon nucleic acid probe used to measuring gene expression  
XX in human breast.  
XX Claim 27; SEQ ID NO 10117; 120pp; English.  
XX The present invention relates to novel single exon nucleic acid probes  
XX (see AA:00010-AAA110067). The present sequence is a peptide encoded by one  
XX such probe. The probes are useful for measuring human gene expression in  
XX a human breast sample, where the probe hybridizes at high stringency to a  
XX nucleic acid expressed in the human breast. The probes are useful for  
XX predicting, diagnosing, grading, staging, monitoring and prognosing  
XX diseases of the human breast, particularly those diseases with polygenic  
XX aetiology. The diseases include breast cancer, disorders of development,  
XX and immediately diseases of the breast, fibrocystic changes, proliferative  
XX breast disease and non carcinoma tumors.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WPI  
XX at <http://wipo.int/pub/publist/pat sequences>.  
XX Sequence 65 AA.  
Query Match  
Best local similarity 100.0%, Identical 100.0%, 100.0%  
Matches 65; Conservative 65; Variable 0; Gaps 0;  
CY 114 DYSELENE-ORNDALVORHLETCRNGJNKHFNCHGHEHLETCOSLAYEQGANIPAA 173  
SD 1 DYSELENE-ORNDALVORHLETCRNGJNKHFNCHGHEHLETCOSLAYEQGANIPAA 60  
QY 174 ELKPT 178  
IIII  
FY 1 ELKPT 65  
RESULT 13  
AB035191  
ID AB035191; standard; Peptide; 65 AA.  
XX AB035191;  
XX  
XX 19 AUG 2002 (first entry)  
XX Human peptide encoded by genome derived single exon probe SEQ ID 25056.  
XX Human single exon probe; asthma; lung cancer; COPD; ILD;  
XX chronic obstructive pulmonary disease; interstitial lung disease;  
XX familial idiopathic pulmonary fibrosis; neurofibromatosis;  
XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
XX Refractory-Pudlak syndrome; sarcoidosis; pulmonary histiocytosis;  
XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karagenen syndrome;

KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
KW primary ciliary dyskinesia; pulmonary hypertension;  
XX hyaline membrane disease.  
XX Homo sapiens.  
XX W02001:56001 A2.  
XX 15-NOV-2001.  
XX 30-JAN-2001; 2001WO-US00465.  
XX 04-FEB-2000; 2000US-1803129.  
XX 26-MAY-2000; 2000US-2074569.  
XX 30-JUN-2000; 2000US-2609408.  
XX 03-AUG-2000; 2000US-2632166.  
XX 21-SEP-2000; 2000US-2146679.  
XX 27-SEP-2000; 2000US-2163599.  
XX 04-OCT-2000; 2000US-0004263.  
XX MOLECULAR DYNAMICS INC.  
XX Peen SG, Hanzel DK, Chen W, Rank DS;  
XX WPI; 2002:114183/15.  
XX Spatially-addressable set of single exon nucleic acid probes, used to  
XX measure gene expression in human lung samples.  
XX Claim 27; SEQ ID NO 25056; 634pp; English.  
XX The invention relates to a spatially-addressable set of single exon  
XX nucleic acid probes for measuring gene expression in a sample derived  
XX from human lung comprising single exon nucleic acid probes having one of  
XX 12614 nucleic acid sequences mentioned in the specification, or their  
XX complements or the 12387 open reading frames derived from the 12614  
XX probes. Also included are a microarray comprising the novel set of  
XX probes; the novel set of probes which hybridise at high stringency to a  
XX nucleic acid expressed in the human lung; measuring gene expression in a  
XX sample derived from human lung, comprising (a) contacting the array with  
XX a collection of detectably labeled nucleic acids derived from human lung  
XX mRNA, and (b) measuring the label detectably bound to each probe of  
XX the array; identifying exons in a eukaryotic genome, comprising  
XX (a) algorithmically predicting at least one exon from genomic sequences  
XX of the eukaryote; and (b) detecting specific hybridisation of detectably  
XX labeled nucleic acids from eukaryotic lung mRNA, to a single exon probe,  
XX having a fragment identical to the predicted exon, the probe is included  
XX in the above mentioned microarray; assigning exons to a single gene,  
XX comprising (a) identifying exons from genomic sequence by the method  
XX above and (b) measuring the expression of each of the exons in several  
XX tissues and/or cell types using hybridisation to a single exon  
XX microarrays having a probe with the exon, where a common pattern of  
XX expression of the exons in the tissues and/or cell types indicates that  
XX the exons should be assigned to a single gene; a peptide comprising one  
XX of 12614 sequences mentioned in the specification, or encoded by the  
XX probes/open reading frames (ORFs). The probes are used for gene  
XX expression analysis, and for identifying exons in a gene, particularly  
XX using human lung derived mRNA and for the study of lung diseases  
XX such as asthma, lung cancer, chronic obstructive pulmonary disease  
XX (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
XX fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,  
XX Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
XX histiocytosis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
XX pulmonary alveolar proteinosis, Karagenen syndrome, fibrocystic  
XX pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
XX and hyaline membrane disease. The present sequence is a peptide/protein  
XX encoded by a single exon probe of the invention.  
XX Note: The sequence data for this patent did not form part  
XX of the printed specification, but was obtained in electronic  
XX format directly from WPI at  
XX <http://wipo.int/pub/publist/pat sequences>.  
XX Sequence 65 AA;





DE 11 UNFFFEAFOSULLTDTQFENRDNKKEHLEIVOKTINFEIDVASOMAEFLVZELYS 70  
QY 104 VOFTEQVKEOVSEHFNSELOPEKALNOKHUTGCHWOCVLEENV CHKEPADIQSSLA 162  
DE 71 TLKPYMLKHEHQUSIEELAEZELLOSHNRIEWEACTDIOGQSHSEPTD PLOS-G 128  
QY 163 YLEQASANIP ALEKP 177  
DE 109 MLQUGKEMPPVNGUTFFRP 147

Search completed: October 20, 2003, 11:34:37  
Web time: 1.90 secs



Matches 126; Conservative 0; Mismatches 113; Indels 4; Gaps 1;  
 Y 644 TATGAGTTACTTTAGTTTATGCTCCCATGAAATTTCACTATTTTATAGCT 703  
 b 15947 TATTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 15868  
 Y 704 GTTAATTTTGTAGTACTTTATACATGCTGTAGCTTGATTAACAGTAAGT 763  
 b 15887 TTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 15828  
 Y 764 TTTTGTGTATTTACCAAGTTAGACTGTGATATGATGACAGATTCCTTTT 823  
 b 15827 TTTTGTGTATTTAATAAATTTTATTTTATTTTATTTTATTTTATTT 15772  
 Y 824 GCGTGTATTTTATTAATTTTGCATGCTTTTATGCTTTTATGCTTTT 883  
 b 15771 TTTTGTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 15712  
 Y 884 TTT 886  
 b 15711 TTT 15709

ESUIT 2  
 S-08-232-463 14/7  
 Sequence 14, Application US/06232463  
 Patent No. 5670367  
 GENERAL INFORMATION:  
 APPLICANT: DORNER, F.  
 APPLICANT: SCHUEFLINGER, F.  
 APPLICANT: FALKNER, F.G.  
 TITLE OF INVENTION: RECOMBINANT FOXLPOX VIRUS  
 NUMBER OF SEQUENCES: 52  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: POLNY & Lardner  
 STREET: 1800 Playona Road, Suite 500  
 CITY: Alexandria  
 STATE: VA  
 COUNTRY: USA  
 ZIP: 22304-0244  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: F-505 disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/28/232,463  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/07/935,313  
 FILING DATE:  
 APPLICATION NUMBER: EP 91 114 300.6  
 FILING DATE: 76-AUG-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: BENT, Stephen A.  
 REGISTRATION NUMBER: 29,768  
 REFERENCE/DOCKET NUMBER: 30472/214 INXJ  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 1703/636-9300  
 TELEFAX: 1703/636-4109  
 TELEX: 894149

INFORMATION FOR SEQ ID NO: 14:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 7218 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 IMMEDIATE SOURCE:  
 CLONE: pTZ9pc-F15  
 US-08-232-463-14

Query Match 3.4%; Score 43.4; DB 1; Length 7218;

Best Local Similarity 10.1%; P-adj. No. 0.074;  
 Matches 88; Conservative 174; Mismatches 16%; Indels 0; Gaps 0;  
 QY 212 TCAGGAATAAATTCANAGCTGTATCAATTAATCAAGATTTGTGATATTCGAAG 171  
 CB 1515 TTAATTTCAAAAGAAATATGATATATATATATATATATATATATATATATAT 1456  
 QY 222 AATACACAAATTTTTCACAAATAATTTACAGTATTTTCAGAAACAGATTA 331  
 CB 1455 AATATAGAAATTTTACATATATATATATATATATATATATATATATATAT 1196  
 QY 332 AATATCAAGAAAT 391  
 CB 1395 AATATCAAGAAAT 1336  
 QY 392 ATAGAAAT 451  
 CB 1345 AATATCAAGAAAT 1276  
 QY 442 ATATCAAGAAAT 511  
 CB 1294 AATATCAAGAAAT 1216  
 QY 512 AATATCAAGAAAT 571  
 CB 1215 AATATCAAGAAAT 1156  
 QY 572 ATATCAAGAAAT 586  
 CB 1194 AATATCAAGAAAT 1134

RESULT 3  
 S-071843-6  
 Patent No. 5171843  
 APPLICANT: KUSFERNBERGER, VIKTOR  
 TITLE OF INVENTION: IMMUNOGENIC POLYPEPTIDES AND METHOD FOR  
 PURIFYING IT  
 NUMBER OF SEQUENCES: 13  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/184,316  
 FILING DATE: 15-MAR-1996  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 154,140  
 FILING DATE: 15-MAR-1996  
 APPLICATION NUMBER: 115,634  
 FILING DATE: 26-OCT-1997  
 APPLICATION NUMBER: 646,993  
 FILING DATE: 16-SEP-1999  
 SEQ ID NO:18  
 LENGTH: 117  
 S-071843-6

Query Match 3.3%; Score 42.4; DB 6; Length 1167;  
 Best Local Similarity 66.1%; P-adj. No. 0.031;  
 Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 318 CAGAAAAGAGAGAAATATATATATATATATATATATATATATATATATAT 477  
 CB 447 CAGAGAGAGAGAAATATATATATATATATATATATATATATATATATAT 458  
 QY 376 AAAAT 487  
 CB 509 AAT 478  
 QY 438 CAGATCAATTTTCAATATATATATATATATATATATATATATATATATAT 454  
 CB 569 CAGATCAATTTTCAATATATATATATATATATATATATATATATATATAT 549

RESULT 4  
 US-09-414-010-3  
 Sequence 3, Application US/09414010

TELEPHONE: 619-235-6000  
TELEFAX: 619-235-0170  
INFORMATION FOR SEQ ID NO: 137  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19124 base pairs  
TYPE: nucleic acid  
STANDARDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTISENSE: NO  
US-08-497-849B.13

Query Match: 3.38; Score 42.2; DB 2; Length 19124;  
Best Local Similarity: 54.31; Prod. No. 0.29;  
Matches 128; Conservative 0; Mismatches 88; Indels 3; Gaps 1;

QY 677 AAAAAATTCACATATTTTATAGCTGTTAAATCTTGGAGTACTTTTACAAATAGTCTGT 736  
DB 17619 ATATATCTTGATATATGATATGTTTATTTTATTTATTTATTTATTTATTTATTT 17698

QY 737 AACTGATMAAACCAATTAATATTTTTCGCTTAGCAAAAGTTAGACTTGAAT 796  
DB 17699 TTCTGGATA...CTTGTAAAGAAATGTTTGTGTATATCATATATATATTTTAA 17755

QY 797 AAGATGACACAGATGTTTCTTATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 856  
DB 17756 ATAAATTTACAGATATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 17815

QY 857 TCATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 879  
DB 17816 ATATATATATATATATTTT 1814

RESULT 6  
US-09-601-198-117C  
Sequence: V, Application US/09601198  
Patent No. 651543  
GENERAL INFORMATION:  
APPLICANT: Cassell, Gail H.  
APPLICANT: Green, Elissa Y.  
APPLICANT: Glass, Jennifer S.  
APPLICANT: Glass, John D.  
APPLICANT: Heider, Cheryl P.  
APPLICANT: Iwakura, Eiji  
TITLE OF INVENTION: NUCLEOTIDE SEQUENCES AND METHODS FOR DETECTING HEPATITIS  
FILE NUMBER: CAP 1345132  
CURRENT APPLICATION NUMBER: US/09601198  
CURRENT FILING DATE: 2000-12-04  
PRIOR APPLICATION NUMBER: 60/031189  
PRIOR FILING DATE: 1998-01-12  
NUMBER OF SEQ ID NOS: 181  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO: 1  
TYPE: cDNA  
ORGANISM: Ureaplasma urealyticum  
US-09-601-198-11

Query Match: 3.28; Score 41.6; DB 4; Length 810;  
Best Local Similarity: 49.18; Prod. No. 0.063;  
Matches 110; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 647 TGAGTTCATTTTATGTTTATGCTCCCAATGAAAATTTCCACTATTTTATAGCTTT 706  
DB 241 TAAGTTCGCACTTTTATTTTATTTCTATAGCAATCTCTCAATTCCTGTAAGTTTAA 182

QY 707 AATTTCTTGAGTACTTTATAGATGTTCTGTAGCTGATAAACCAGTAAGTATTTT 766  
DB 181 ATCAATTAAGGATTTATTAATTCATGASAAATTAATCATCATGACACCATATATAT 122

QY 767 TTGTCTTTAGCAAGTTTATTTCTGTAATATGATGACACAGATTTCTTTTATTTGCTGCT 826

TELEPHONE: 619-235-6000  
TELEFAX: 619-235-0170  
INFORMATION FOR SEQ ID NO: 137  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19124 base pairs  
TYPE: nucleic acid  
STANDARDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTISENSE: NO  
US-08-497-849B.13

Query Match: 3.38; Score 42.2; DB 2; Length 19124;  
Best Local Similarity: 54.31; Prod. No. 0.29;  
Matches 128; Conservative 0; Mismatches 88; Indels 3; Gaps 1;

QY 677 AAAAAATTCACATATTTTATAGCTGTTAAATCTTGGAGTACTTTTACAAATAGTCTGT 736  
DB 17619 ATATATCTTGATATATGATATGTTTATTTTATTTATTTATTTATTTATTTATTT 17698

QY 737 AACTGATMAAACCAATTAATATTTTTCGCTTAGCAAAAGTTAGACTTGAAT 796  
DB 17699 TTCTGGATA...CTTGTAAAGAAATGTTTGTGTATATCATATATATATTTTAA 17755

QY 797 AAGATGACACAGATGTTTCTTATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 856  
DB 17756 ATAAATTTACAGATATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 17815

QY 857 TCATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 879  
DB 17816 ATATATATATATATTTT 1814

RESULT 6  
US-09-601-198-117C  
Sequence: V, Application US/09601198  
Patent No. 651543  
GENERAL INFORMATION:  
APPLICANT: Cassell, Gail H.  
APPLICANT: Green, Elissa Y.  
APPLICANT: Glass, Jennifer S.  
APPLICANT: Glass, John D.  
APPLICANT: Heider, Cheryl P.  
APPLICANT: Iwakura, Eiji  
TITLE OF INVENTION: NUCLEOTIDE SEQUENCES AND METHODS FOR DETECTING HEPATITIS  
FILE NUMBER: CAP 1345132  
CURRENT APPLICATION NUMBER: US/09601198  
CURRENT FILING DATE: 2000-12-04  
PRIOR APPLICATION NUMBER: 60/031189  
PRIOR FILING DATE: 1998-01-12  
NUMBER OF SEQ ID NOS: 181  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO: 1  
TYPE: cDNA  
ORGANISM: Ureaplasma urealyticum  
US-09-601-198-11

Query Match: 3.28; Score 41.6; DB 4; Length 810;  
Best Local Similarity: 49.18; Prod. No. 0.063;  
Matches 110; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 647 TGAGTTCATTTTATGTTTATGCTCCCAATGAAAATTTCCACTATTTTATAGCTTT 706  
DB 241 TAAGTTCGCACTTTTATTTTATTTCTATAGCAATCTCTCAATTCCTGTAAGTTTAA 182

QY 707 AATTTCTTGAGTACTTTATAGATGTTCTGTAGCTGATAAACCAGTAAGTATTTT 766  
DB 181 ATCAATTAAGGATTTATTAATTCATGASAAATTAATCATCATGACACCATATATAT 122

QY 767 TTGTCTTTAGCAAGTTTATTTCTGTAATATGATGACACAGATTTCTTTTATTTGCTGCT 826

























Family	No.	Query		Length	P#	ID	Description
		Score	Match				
1	1	12.5	11.2	834	1	7474	HUMAN
2	1	12.5	10.6	983	1	421	MUSE
3	4	96.5	10.4	340	1	1095	BRUSH
4	4	96.5	10.4	1112	1	1095	HUMAN
5	6	95	10.3	454	1	1095	HUMAN
6	6	94	10.2	528	1	1095	YEAST
7	7	97	10.1	1127	1	1095	HUMAN
8	4	94.5	10.1	922	1	1095	HUMAN
9	8	92	10.0	828	1	1095	HUMAN
10	9	91	9.9	5147	1	1095	HUMAN
11	11	88	9.8	388	1	1095	HUMAN
12	12	90	9.8	1223	1	1095	YEAST
13	13	89.5	9.7	1360	1	1095	HUMAN
14	14	90.5	9.7	1361	1	1095	HUMAN
15	15	89	9.7	373	1	1095	HUMAN
16	16	89	9.7	1353	1	1095	HUMAN
17	17	89	9.7	5338	1	1095	HUMAN
18	18	89.5	9.6	910	1	1095	HUMAN
19	19	89.5	9.5	1022	1	1095	HUMAN
20	20	87	9.5	344	1	1095	HUMAN
21	21	87	9.5	773	1	1095	HUMAN
22	22	87	9.5	3018	1	1095	HUMAN
23	23	86.5	9.4	757	1	1095	HUMAN
24	24	86.5	9.4	1021	1	1095	HUMAN
25	25	86	9.4	209	1	1095	HUMAN
26	26	86	9.4	680	1	1095	HUMAN
27	27	85.5	9.3	408	1	1095	HUMAN
28	28	85.5	9.3	961	1	1095	HUMAN
29	29	85	9.3	131	1	1095	HUMAN
30	30	85	9.3	621	1	1095	YEAST
31	31	85	9.3	631	1	1095	YEAST
32	32	84.5	9.2	1527	1	1095	YEAST
33	33	84.5	9.2	1359	1	1095	YEAST







```

VARSPLIC 678 999 /FTID=VSP 005581.
Missing (in isoform A).
/FTID=VSP 005582.
H -> C (IN REF. 3).
CONFLICT 391 391 H -> C (IN REF. 3).
CONFLICT 401 401 H -> F (IN REF. 3).
CONFLICT 407 407 V -> SQEQYQEQSL (IN REF. 3).
CONFLICT 698 699 DK -> EQ (IN REF. 3).
CONFLICT 965 999 AEFETSAAYAVPACGELTPDEVAKLAANGI -> PV
NTRPPPKYLPKWLAR (IN REF. 3).
Q SEQUENCE 999 AA: 111551 MW: 23665944 AC070662 CRC64;
Query Watch: 10.4%; Score 95.5; DB 1; Length: 999;
Best Local Similarity: 23.4%; Pred. No. 13;
Matches 41; Conservative 30; Mismatches 8; Indels 23; Gaps 5;
Y 16 PQAPFGLGQASLIQACAPRPSSTLVDSLESPACFASIVSQDYNGVGTOR-EIR 74
b 449 PNCQQLPAQSPQCAIDHLEPHTS...DRPEPMFAADAHAEAEARLRSEQKREID 504
Y 75 TGVOOTQKRF--LDIARQEC-----FEQKMGICQVQKPEGVKEDVSELNRE 121
b 505 TIFDKVRFLELDLQCOLEPRLIARDKLTIRKMSPIETGVVIEPSTACADYLSALP 564
Y 122 LQRIALVAVETKLRHWQVLEIDNVKKKACITQSLAYLEQASNIAPLX 176
b 565 ISKADKTIQELTVAELDELYTDFAVQCRQTYKNITQNELAGLOGR- APLQ 674

```

```

RESULT 4
ER3_HUMAN STANDARD; PRT; 1210 AA.
P56645; Q969K6; Q96S77; Q96S78; Q96C03; Q9NSH9; G8G0D8;
15-JUL-1999 Ref. 38, Created;
16-OCT-2001 Ref. 42, Last sequence update;
15-SEP-2003 Ref. 42, Last annotation update;
Period circadian protein 3 (PER3).
S Homo sapiens (Human);
P Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID:9606;
(1)
SEQUENCE FROM N.A.
TISSUE=Brain;
Nagase T., Kikuna K., Ohara O.,
Submitted (JUL 2003) to the EMBL/GenBank/DBJ databases.
(2)
SEQUENCE FROM N.A.
Knoges S., Buckle E.,
Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
(3)
SEQUENCE OF 44 1210 FROM N.A.
Pearce A., (DEC-1999); to the EMBL/GenBank/DBJ databases.
(4)
SEQUENCE OF 114-119 FROM N.A.
Mistry S.,
Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
(5)
SEQUENCE OF 44-91 AND 732-890 FROM N.A., AND VARIANT ALA 864.
Edisawa T., Chihara M., Kajimura K., Mishima K., Kamei Y., Katon M.,
Watanabe T., Sakimoto M., Shibui K., Kim K., Kudo Y., Czeiz Y.,
Iroue Y., Yamada N., Nagase T., Ozaki K., Ohara O., Ishida N.,
Okawa M., Takahashi K.;
"Association of structural polymorphisms in human period3 gene with
delayed sleep phase syndrome.";
Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
(6)
FUNCTION: CIRCADIAN REGULATOR THAT MAY ACT AS A TRANSCRIPTION
LOOP. BEHAVES AS A NEGATIVE ELEMENT IN CIRCADIAN TRANSCRIPTIONAL
LOOP. DOES NOT APPEAR TO BIND DNA, SUGGESTING INDIRECT
TRANSCRIPTIONAL INHIBITION. THE EXPRESSION RHYTHMS APPEAR TO
ORIGINATE FROM RETINA (BY SIMILARITY).
(7)
SUBCELLULAR LOCATION: Nuclear (potential).

```

```

-1- SIMILARITY: BELONGS TO THE BASIC HELIX LOOP-HELIX (BHLH) FAMILY OF
TRANSCRIPTION FACTORS.
-1- SIMILARITY: Contains 2 PAS (PER-ARNT-SIM) dimerization domains.
-1- SIMILARITY: Contains 1 PAS-associated C-terminal (PAC) domain.
CC THIS SWISS PROT entry is copyrighted. It is published through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announcement/
or send an email to license@sib.ch).
DR ENML; AB047636; BAB12345;
DR ENML; A1573984; CAU76344;
DR ENML; 2986844; CAB631147;
DR ENML; AB047512; BAB12345;
DR ENML; AB047510; BAB12345;
DR ENML; AB047511; BAB12345;
DR ENML; AB047513; BAB12345;
DR ENML; AB047514; BAB12345;
DR ENML; AB047515; BAB12345;
DR ENML; AB047516; BAB12345;
DR ENML; AB047517; BAB12345;
DR ENML; AB047518; BAB12345;
DR ENML; AB047519; BAB12345;
DR ENML; AB047520; BAB12345;
DR ENML; AB047521; BAB12345;
DR ENML; AB047522; BAB12345;
DR ENML; AB047523; BAB12345;
DR ENML; AB047524; BAB12345;
DR ENML; AB047525; BAB12345;
DR ENML; AB047526; BAB12345;
DR ENML; AB047527; BAB12345;
DR ENML; AB047528; BAB12345;
DR ENML; AB047529; BAB12345;
DR ENML; AB047530; BAB12345;
DR ENML; AB047531; BAB12345;
DR ENML; AB047532; BAB12345;
DR ENML; AB047533; BAB12345;
DR ENML; AB047534; BAB12345;
DR ENML; AB047535; BAB12345;
DR ENML; AB047536; BAB12345;
DR ENML; AB047537; BAB12345;
DR ENML; AB047538; BAB12345;
DR ENML; AB047539; BAB12345;
DR ENML; AB047540; BAB12345;
DR ENML; AB047541; BAB12345;
DR ENML; AB047542; BAB12345;
DR ENML; AB047543; BAB12345;
DR ENML; AB047544; BAB12345;
DR ENML; AB047545; BAB12345;
DR ENML; AB047546; BAB12345;
DR ENML; AB047547; BAB12345;
DR ENML; AB047548; BAB12345;
DR ENML; AB047549; BAB12345;
DR ENML; AB047550; BAB12345;
DR ENML; AB047551; BAB12345;
DR ENML; AB047552; BAB12345;
DR ENML; AB047553; BAB12345;
DR ENML; AB047554; BAB12345;
DR ENML; AB047555; BAB12345;
DR ENML; AB047556; BAB12345;
DR ENML; AB047557; BAB12345;
DR ENML; AB047558; BAB12345;
DR ENML; AB047559; BAB12345;
DR ENML; AB047560; BAB12345;
DR ENML; AB047561; BAB12345;
DR ENML; AB047562; BAB12345;
DR ENML; AB047563; BAB12345;
DR ENML; AB047564; BAB12345;
DR ENML; AB047565; BAB12345;
DR ENML; AB047566; BAB12345;
DR ENML; AB047567; BAB12345;
DR ENML; AB047568; BAB12345;
DR ENML; AB047569; BAB12345;
DR ENML; AB047570; BAB12345;
DR ENML; AB047571; BAB12345;
DR ENML; AB047572; BAB12345;
DR ENML; AB047573; BAB12345;
DR ENML; AB047574; BAB12345;
DR ENML; AB047575; BAB12345;
DR ENML; AB047576; BAB12345;
DR ENML; AB047577; BAB12345;
DR ENML; AB047578; BAB12345;
DR ENML; AB047579; BAB12345;
DR ENML; AB047580; BAB12345;
DR ENML; AB047581; BAB12345;
DR ENML; AB047582; BAB12345;
DR ENML; AB047583; BAB12345;
DR ENML; AB047584; BAB12345;
DR ENML; AB047585; BAB12345;
DR ENML; AB047586; BAB12345;
DR ENML; AB047587; BAB12345;
DR ENML; AB047588; BAB12345;
DR ENML; AB047589; BAB12345;
DR ENML; AB047590; BAB12345;
DR ENML; AB047591; BAB12345;
DR ENML; AB047592; BAB12345;
DR ENML; AB047593; BAB12345;
DR ENML; AB047594; BAB12345;
DR ENML; AB047595; BAB12345;
DR ENML; AB047596; BAB12345;
DR ENML; AB047597; BAB12345;
DR ENML; AB047598; BAB12345;
DR ENML; AB047599; BAB12345;
DR ENML; AB047600; BAB12345;
DR ENML; AB047601; BAB12345;
DR ENML; AB047602; BAB12345;
DR ENML; AB047603; BAB12345;
DR ENML; AB047604; BAB12345;
DR ENML; AB047605; BAB12345;
DR ENML; AB047606; BAB12345;
DR ENML; AB047607; BAB12345;
DR ENML; AB047608; BAB12345;
DR ENML; AB047609; BAB12345;
DR ENML; AB047610; BAB12345;
DR ENML; AB047611; BAB12345;
DR ENML; AB047612; BAB12345;
DR ENML; AB047613; BAB12345;
DR ENML; AB047614; BAB12345;
DR ENML; AB047615; BAB12345;
DR ENML; AB047616; BAB12345;
DR ENML; AB047617; BAB12345;
DR ENML; AB047618; BAB12345;
DR ENML; AB047619; BAB12345;
DR ENML; AB047620; BAB12345;
DR ENML; AB047621; BAB12345;
DR ENML; AB047622; BAB12345;
DR ENML; AB047623; BAB12345;
DR ENML; AB047624; BAB12345;
DR ENML; AB047625; BAB12345;
DR ENML; AB047626; BAB12345;
DR ENML; AB047627; BAB12345;
DR ENML; AB047628; BAB12345;
DR ENML; AB047629; BAB12345;
DR ENML; AB047630; BAB12345;
DR ENML; AB047631; BAB12345;
DR ENML; AB047632; BAB12345;
DR ENML; AB047633; BAB12345;
DR ENML; AB047634; BAB12345;
DR ENML; AB047635; BAB12345;
DR ENML; AB047636; BAB12345;
DR ENML; AB047637; BAB12345;
DR ENML; AB047638; BAB12345;
DR ENML; AB047639; BAB12345;
DR ENML; AB047640; BAB12345;
DR ENML; AB047641; BAB12345;
DR ENML; AB047642; BAB12345;
DR ENML; AB047643; BAB12345;
DR ENML; AB047644; BAB12345;
DR ENML; AB047645; BAB12345;
DR ENML; AB047646; BAB12345;
DR ENML; AB047647; BAB12345;
DR ENML; AB047648; BAB12345;
DR ENML; AB047649; BAB12345;
DR ENML; AB047650; BAB12345;
DR ENML; AB047651; BAB12345;
DR ENML; AB047652; BAB12345;
DR ENML; AB047653; BAB12345;
DR ENML; AB047654; BAB12345;
DR ENML; AB047655; BAB12345;
DR ENML; AB047656; BAB12345;
DR ENML; AB047657; BAB12345;
DR ENML; AB047658; BAB12345;
DR ENML; AB047659; BAB12345;
DR ENML; AB047660; BAB12345;
DR ENML; AB047661; BAB12345;
DR ENML; AB047662; BAB12345;
DR ENML; AB047663; BAB12345;
DR ENML; AB047664; BAB12345;
DR ENML; AB047665; BAB12345;
DR ENML; AB047666; BAB12345;
DR ENML; AB047667; BAB12345;
DR ENML; AB047668; BAB12345;
DR ENML; AB047669; BAB12345;
DR ENML; AB047670; BAB12345;
DR ENML; AB047671; BAB12345;
DR ENML; AB047672; BAB12345;
DR ENML; AB047673; BAB12345;
DR ENML; AB047674; BAB12345;
DR ENML; AB047675; BAB12345;
DR ENML; AB047676; BAB12345;
DR ENML; AB047677; BAB12345;
DR ENML; AB047678; BAB12345;
DR ENML; AB047679; BAB12345;
DR ENML; AB047680; BAB12345;
DR ENML; AB047681; BAB12345;
DR ENML; AB047682; BAB12345;
DR ENML; AB047683; BAB12345;
DR ENML; AB047684; BAB12345;
DR ENML; AB047685; BAB12345;
DR ENML; AB047686; BAB12345;
DR ENML; AB047687; BAB12345;
DR ENML; AB047688; BAB12345;
DR ENML; AB047689; BAB12345;
DR ENML; AB047690; BAB12345;
DR ENML; AB047691; BAB12345;
DR ENML; AB047692; BAB12345;
DR ENML; AB047693; BAB12345;
DR ENML; AB047694; BAB12345;
DR ENML; AB047695; BAB12345;
DR ENML; AB047696; BAB12345;
DR ENML; AB047697; BAB12345;
DR ENML; AB047698; BAB12345;
DR ENML; AB047699; BAB12345;
DR ENML; AB047700; BAB12345;
DR ENML; AB047701; BAB12345;
DR ENML; AB047702; BAB12345;
DR ENML; AB047703; BAB12345;
DR ENML; AB047704; BAB12345;
DR ENML; AB047705; BAB12345;
DR ENML; AB047706; BAB12345;
DR ENML; AB047707; BAB12345;
DR ENML; AB047708; BAB12345;
DR ENML; AB047709; BAB12345;
DR ENML; AB047710; BAB12345;
DR ENML; AB047711; BAB12345;
DR ENML; AB047712; BAB12345;
DR ENML; AB047713; BAB12345;
DR ENML; AB047714; BAB12345;
DR ENML; AB047715; BAB12345;
DR ENML; AB047716; BAB12345;
DR ENML; AB047717; BAB12345;
DR ENML; AB047718; BAB12345;
DR ENML; AB047719; BAB12345;
DR ENML; AB047720; BAB12345;
DR ENML; AB047721; BAB12345;
DR ENML; AB047722; BAB12345;
DR ENML; AB047723; BAB12345;
DR ENML; AB047724; BAB12345;
DR ENML; AB047725; BAB12345;
DR ENML; AB047726; BAB12345;
DR ENML; AB047727; BAB12345;
DR ENML; AB047728; BAB12345;
DR ENML; AB047729; BAB12345;
DR ENML; AB047730; BAB12345;
DR ENML; AB047731; BAB12345;
DR ENML; AB047732; BAB12345;
DR ENML; AB047733; BAB12345;
DR ENML; AB047734; BAB12345;
DR ENML; AB047735; BAB12345;
DR ENML; AB047736; BAB12345;
DR ENML; AB047737; BAB12345;
DR ENML; AB047738; BAB12345;
DR ENML; AB047739; BAB12345;
DR ENML; AB047740; BAB12345;
DR ENML; AB047741; BAB12345;
DR ENML; AB047742; BAB12345;
DR ENML; AB047743; BAB12345;
DR ENML; AB047744; BAB12345;
DR ENML; AB047745; BAB12345;
DR ENML; AB047746; BAB12345;
DR ENML; AB047747; BAB12345;
DR ENML; AB047748; BAB12345;
DR ENML; AB047749; BAB12345;
DR ENML; AB047750; BAB12345;
DR ENML; AB047751; BAB12345;
DR ENML; AB047752; BAB12345;
DR ENML; AB047753; BAB12345;
DR ENML; AB047754; BAB12345;
DR ENML; AB047755; BAB12345;
DR ENML; AB047756; BAB12345;
DR ENML; AB047757; BAB12345;
DR ENML; AB047758; BAB12345;
DR ENML; AB047759; BAB12345;
DR ENML; AB047760; BAB12345;
DR ENML; AB047761; BAB12345;
DR ENML; AB047762; BAB12345;
DR ENML; AB047763; BAB12345;
DR ENML; AB047764; BAB12345;
DR ENML; AB047765; BAB12345;
DR ENML; AB047766; BAB12345;
DR ENML; AB047767; BAB12345;
DR ENML; AB047768; BAB12345;
DR ENML; AB047769; BAB12345;
DR ENML; AB047770; BAB12345;
DR ENML; AB047771; BAB12345;
DR ENML; AB047772; BAB12345;
DR ENML; AB047773; BAB12345;
DR ENML; AB047774; BAB12345;
DR ENML; AB047775; BAB12345;
DR ENML; AB047776; BAB12345;
DR ENML; AB047777; BAB12345;
DR ENML; AB047778; BAB12345;
DR ENML; AB047779; BAB12345;
DR ENML; AB047780; BAB12345;
DR ENML; AB047781; BAB12345;
DR ENML; AB047782; BAB12345;
DR ENML; AB047783; BAB12345;
DR ENML; AB047784; BAB12345;
DR ENML; AB047785; BAB12345;
DR ENML; AB047786; BAB12345;
DR ENML; AB047787; BAB12345;
DR ENML; AB047788; BAB12345;
DR ENML; AB047789; BAB12345;
DR ENML; AB047790; BAB12345;
DR ENML; AB047791; BAB12345;
DR ENML; AB047792; BAB12345;
DR ENML; AB047793; BAB12345;
DR ENML; AB047794; BAB12345;
DR ENML; AB047795; BAB12345;
DR ENML; AB047796; BAB12345;
DR ENML; AB047797; BAB12345;
DR ENML; AB047798; BAB12345;
DR ENML; AB047799; BAB12345;
DR ENML; AB047800; BAB12345;
DR ENML; AB047801; BAB12345;
DR ENML; AB047802; BAB12345;
DR ENML; AB047803; BAB12345;
DR ENML; AB047804; BAB12345;
DR ENML; AB047805; BAB12345;
DR ENML; AB047806; BAB12345;
DR ENML; AB047807; BAB12345;
DR ENML; AB047808; BAB12345;
DR ENML; AB047809; BAB12345;
DR ENML; AB047810; BAB12345;
DR ENML; AB047811; BAB12345;
DR ENML; AB047812; BAB12345;
DR ENML; AB047813; BAB12345;
DR ENML; AB047814; BAB12345;
DR ENML; AB047815; BAB12345;
DR ENML; AB047816; BAB12345;
DR ENML; AB047817; BAB12345;
DR ENML; AB047818; BAB12345;
DR ENML; AB047819; BAB12345;
DR ENML; AB047820; BAB12345;
DR ENML; AB047821; BAB12345;
DR ENML; AB047822; BAB12345;
DR ENML; AB047823; BAB12345;
DR ENML; AB047824; BAB12345;
DR ENML; AB047825; BAB12345;
DR ENML; AB047826; BAB12345;
DR ENML; AB047827; BAB12345;
DR ENML; AB047828; BAB12345;
DR ENML; AB047829; BAB12345;
DR ENML; AB047830; BAB12345;
DR ENML; AB047831; BAB12345;
DR ENML; AB047832; BAB12345;
DR ENML; AB047833; BAB12345;
DR ENML; AB047834; BAB12345;
DR ENML; AB047835; BAB12345;
DR ENML; AB047836; BAB12345;
DR ENML; AB047837; BAB12345;
DR ENML; AB047838; BAB12345;
DR ENML; AB047839; BAB12345;
DR ENML; AB047840; BAB12345;
DR ENML; AB047841; BAB12345;
DR ENML; AB047842; BAB12345;
DR ENML; AB047843; BAB12345;
DR ENML; AB047844; BAB12345;
DR ENML; AB047845; BAB12345;
DR ENML; AB047846; BAB12345;
DR ENML; AB047847; BAB12345;
DR ENML; AB047848; BAB12345;
DR ENML; AB047849; BAB12345;
DR ENML; AB047850; BAB12345;
DR ENML; AB047851; BAB12345;
DR ENML; AB047852; BAB12345;
DR ENML; AB047853; BAB12345;
DR ENML; AB047854; BAB12345;
DR ENML; AB047855; BAB12345;
DR ENML; AB047856; BAB12345;
DR ENML; AB047857; BAB12345;
DR ENML; AB047858; BAB12345;
DR ENML; AB047859; BAB12345;
DR ENML; AB047860; BAB12345;
DR ENML; AB047861; BAB12345;
DR ENML; AB047862; BAB12345;
DR ENML; AB047863; BAB12345;
DR ENML; AB047864; BAB12345;
DR ENML; AB047865; BAB12345;
DR ENML; AB047866; BAB12345;
DR ENML; AB047867; BAB12345;
DR ENML; AB047868; BAB12345;
DR ENML; AB047869; BAB12345;
DR ENML; AB047870; BAB12345;
DR ENML; AB047871; BAB12345;
DR ENML; AB047872; BAB12345;
DR ENML; AB047873; BAB12345;
DR ENML; AB047874; BAB12345;
DR ENML; AB047875; BAB12345;
DR ENML; AB047876; BAB12345;
DR ENML; AB047877; BAB12345;
DR ENML; AB047878; BAB12345;
DR ENML; AB047879; BAB12345;
DR ENML; AB047880; BAB12345;
DR ENML; AB047881; BAB12345;
DR ENML; AB047882; BAB12345;
DR ENML; AB047883; BAB12345;
DR ENML; AB047884; BAB12345;
DR ENML; AB047885; BAB12345;
DR ENML; AB047886; BAB12345;
DR ENML; AB047887; BAB12345;
DR ENML; AB047888; BAB12345;
DR ENML; AB047889; BAB12345;
DR ENML; AB047890; BAB12345;
DR ENML; AB047891; BAB12345;
DR ENML; AB047892; BAB12345;
DR ENML; AB047893; BAB12345;
DR ENML; AB047894; BAB12345;
DR ENML; AB047895; BAB12345;
DR ENML; AB047896; BAB12345;
DR ENML; AB047897; BAB12345;
DR ENML; AB047898; BAB12345;
DR ENML; AB047899; BAB12345;
DR ENML; AB047900; BAB12345;
DR ENML; AB047901; BAB12345;
DR ENML; AB047902; BAB12345;
DR ENML; AB047903; BAB12345;
DR ENML; AB047904; BAB12345;
DR ENML; AB047905; BAB12345;
DR ENML; AB047906; BAB12345;
DR ENML; AB047907; BAB12345;
DR ENML; AB047908; BAB12345;
DR ENML; AB047909; BAB12345;
DR ENML; AB047910; BAB12345;
DR ENML; AB047911; BAB12345;
DR ENML; AB047912; BAB12345;
DR ENML; AB047913; BAB12345;
DR ENML; AB047914; BAB12345;
DR ENML; AB047915; BAB12345;
DR ENML; AB047916; BAB12345;
DR ENML; AB047917; BAB12345;
DR ENML; AB047918; BAB12345;
DR ENML; AB047919; BAB12345;
DR ENML; AB047920; BAB12345;
DR ENML; AB047921; BAB12345;
DR ENML; AB047922; BAB12345;
DR ENML; AB047923; BAB12345;
DR ENML; AB047924; BAB12345;
DR ENML; AB047925; BAB12345;
DR ENML; AB047926; BAB12345;
DR ENML; AB047927; BAB12345;
DR ENML; AB047928; BAB12345;
DR ENML; AB047929; BAB12345;
DR ENML; AB047930; BAB12345;
DR ENML; AB047931; BAB12345;
DR ENML; AB047932; BAB12345;
DR ENML; AB047933; BAB12345;
DR ENML; AB047934; BAB12345;
DR ENML; AB047935; BAB12345;
DR ENML; AB047936; BAB12345;
DR ENML; AB047937; BAB12345;
DR ENML; AB047938; BAB12345;
DR ENML; AB047939; BAB12345;
DR ENML; AB047940; BAB12345;
DR ENML; AB047941; BAB12345;
DR ENML; AB047942; BAB12345;
DR ENML; AB047943; BAB12345;
DR ENML; AB047944; BAB12345;
DR ENML; AB047945; BAB12345;
DR ENML; AB047946; BAB12345;
DR ENML; AB047947; BAB12345;
DR ENML; AB047948; BAB12345;
DR ENML; AB047949; BAB12345;
DR ENML; AB047950; BAB12345;
DR ENML; AB047951; BAB12345;
DR ENML; AB047952; BAB12345;
DR ENML; AB047953; BAB12345;
DR ENML; AB047954; BAB12345;
DR ENML; AB047955; BAB12345;
DR ENML; AB047956; BAB12345;
DR ENML; AB047957; BAB12345;
DR ENML; AB047958; BAB12345;
DR ENML; AB047959; BAB12345;
DR ENML; AB047960; BAB12345;
DR ENML; AB047961; BAB12345;
DR ENML; AB047962; BAB12345;
DR ENML; AB047963; BAB12345;
DR ENML; AB047964; BAB12345;
DR ENML; AB047965; BAB12345;
DR ENML; AB047966; BAB12345;
DR ENML; AB047967; BAB12345;
DR ENML; AB047968; BAB12345;
DR ENML; AB047969; BAB12345;
DR ENML; AB047970; BAB12345;
DR ENML; AB047971; BAB12345;
DR ENML; AB047972; BAB12345;
DR ENML; AB047973; BAB12345;
DR ENML; AB047974; BAB12345;
DR ENML; AB047975; BAB12345;
DR ENML; AB047976; BAB12345;
DR ENML; AB047977; BAB12345;
DR ENML; AB047978; BAB12345;
DR ENML; AB047979; BAB12345;
DR ENML; AB047980; BAB12345;
DR ENML; AB047981; BAB12345;
DR ENML; AB047982; BAB12345;
DR ENML; AB047983; BAB12345;
DR ENML; AB047984; BAB12345;
DR ENML; AB047985; BAB12345;
DR ENML; AB047986; BAB12345;
DR ENML; AB047987; BAB12345;
DR ENML; AB047988; BAB12345;
DR ENML; AB047989; BAB12345;
DR ENML; AB047990; BAB12345;
DR ENML; AB047991; BAB12345;
DR ENML; AB047992; BAB12345;
DR ENML; AB047993; BAB12345;
DR ENML; AB047994; BAB12345;
DR ENML; AB047995; BAB12345;
DR ENML; AB047996; BAB12345;
DR ENML; AB047997; BAB12345;
DR ENML; AB047998; BAB12345;
DR ENML; AB047999; BAB12345;
DR ENML; AB048000; BAB12345;
DR ENML; AB048001; BAB12345;
DR ENML; AB048002; BAB12345;
DR ENML; AB048003; BAB12345;
DR ENML; AB048004; BAB12345;
DR ENML; AB048005; BAB12345;
DR ENML; AB048006; BAB12345;
DR ENML; AB048007; BAB12345;
DR ENML; AB048008; BAB12345;
DR ENML; AB048009; BAB12345;
DR ENML; AB048010; BAB12345;
DR ENML; AB048011; BAB12345;
DR ENML; AB048012; BAB12345;
DR ENML; AB048013; BAB12345;
DR ENML; AB048014; BAB12345;
DR ENML; AB048015; BAB12345;
DR ENML; AB048016; BAB12345;
DR ENML; AB048017; BAB12345;
DR ENML; AB048018; BAB12345;
DR ENML; AB048019; BAB12345;
DR ENML; AB048020; BAB12345;
DR ENML; AB048021; BAB12345;
DR ENML; AB048022; BAB12345;
DR ENML; AB048023; BAB12345;
DR ENML; AB048024; BAB12345;
DR ENML; AB048025; BAB12345;
DR ENML; AB048026; BAB12345;
DR ENML; AB048027; BAB12345;
DR ENML; AB048028; BAB12345;
DR ENML; AB048029; BAB12345;
DR ENML; AB048030; BAB12345;
DR ENML; AB048031; BAB12345;
DR ENML; AB048032; BAB12345;
DR ENML; AB048033; BAB12345;
DR ENML; AB048034; BAB12345;
DR ENML; AB048035; BAB12345;
DR ENML; AB048036; BAB12345;
DR ENML; AB048037; BAB12345;
DR ENML; AB048038; BAB12345;
DR ENML; AB048039; BAB12345;
DR ENML; AB04
```

Bakaryou; Metazoa; Chordata; Canalia; Vertebrata; Euteleostomi;  
OC Mammalia; Primates; Carnivora; Hominoidea; Homo.  
CC NCPI; TAXID=9606;  
CC {}  
EP SEQUENCE FROM N.A.  
PC TINUS-1499;  
RX MEINUNGWOLFS; PUBM-147748;  
RA Sadel M., Berk P., Elm J.A., Kunkel K., Black T., Kaplan M.,  
RA Hutterer K., Leaman C.J.  
PT Characterization of the mammalian YAP (Yes-associated protein) gene  
PT and its role in defining a novel protein module, the WW domain.";  
RI J. Biol. Chem. 276:14741-14741(1991).  
RK {}  
CC INTERACTION WITH WEE1 AND WBP2.  
RX MEINUNGWOLFS; PUBM-92027A.  
PC Chen H.L., Kinkad A., Kwak S.-J., Ling H., Koepf E., Peterson S.,  
RA Kelly J.W., Sudol M.  
PT "Characterization of the WW domain of human Yes associated protein and  
PT its phosphorylation containing tyrosines".  
RI J. Biol. Chem. 272:17070-17070(1997).  
CC {}  
CC {} SCHUCHT; Placids to the SH domain of the YES kinase binds to WBP1  
CC and WBP2.  
CC {} PTM: PHOSPHORYLATED IN SERINE RESIDUES (BY SIMILARITY).  
CC {} {} SIMILARITY: Contains 1 WW domain.  
CC {}  
CC This SWISS PROT entry is obsolete. It is replaced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC at European Bioinformatics Institute. There are no restrictions on its  
CC use by non profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement. See [http://www.ebi.ac.uk/announce/](http://www.ebi.ac.uk/announcement/)  
CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk).  
CC {}  
DE PMH, XP5607; PAA56672 1.  
DE PIF, AA4364; AS6954.  
DE PQY, LMGQ, ZI-DRC-01.  
DE PRK, IK9D, ZB-DRC-01.  
DE PSV, IK9F, ZB-DRC-01.  
DE SGWW, HGNC:1622; YAP.  
DE NRG, RefSeq:  
DE Q3, OICG5525; Protein kinase activity, TAS.  
DE Inhibitor, IRP05153; BCRK.  
DE IRLIG, IRP01202; WW\_FNF\_WWS.  
DE P4CM, PRC337; WW 1.  
DE SMWT, SM\_2466; WW 1.  
DE PRNLY, PRC113; WW\_FNF\_WWS 1.  
DE PRNLY, IRP502; WW\_FNF\_WWS 1.  
DE PRNLY, IRP502; WW\_FNF\_WWS 1.  
DE DECAN, 1, 294 aa.  
DE SEQUENCE: 434 AA; 49 aa MW; 10000 Da pI; 10000 Da.

```

Query Match 20 33 Score 50 DP 11 Length 4547
Post Local Similarity 25067 Prod No 1514
Matched 46 Conservative 21 Mismatched 69 Index 52 Gaps 10
QY 3 APLKXVSGQPPSHV---QATGPHPL---ASLQCAAGAGAPSSSTLV---DELESSEA 54
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12 APQG---GGGPPSFTGCHTSNDELAPATLNAQVAP---PHQIVHVGJSETLEA 67
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 65 PFASLVSGQVNGTQEEETSV---QCTKPKPLARCTGCTPLKFLQCSVCKPQVIRK 114
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
PB 68 LFNVAVPKXTAN-----VQIVNVP-----KALDSEPKPPEP-KSH 104
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 115 VSEIRNELQRDAVQCHLTKRHWQVLEINVYHKKPATPQSLAYLECSANIPAP 174
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 125 SQCASTDAGTASATTCVPR-----AMSPASQLQVANS-----PGT 141
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 175 LKPT 178
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 142 LKPT 145
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 1

```







[illegible]











GenScan Version 5.1.6  
Copyright 1994-2003 CompuGen Ltd.

on track... nucleic search, using sw method

Run on: October 21, 2003, 11:34:32, Search time 99 seconds

without alignment  
569,504 Million cell updates/sec

File: US-10-029-137-1

Project source: 1281

Sequence: 1 candidate produced (1.00% ... alignment accuracy 1281)

Scoring method: GBS-MSC

Gap: 50bp E-6, Gapext: E-6.8

Searched: 569958 seqs, 22569124 residues

Word size: 3

Total number of hits satisfying chosen parameters: 110956

Minimum hit length: 6

Maximum hit length: 2000 bp

Post-processing: listing first 100 summaries

Database: Insured Patents NA  
1: /zeng2-6/ncdat/127/na/na-NA-00000001  
2: /zeng2-6/ncdat/127/na/na-NA-00000001  
3: /zeng2-6/ncdat/127/na/na-NA-00000001  
4: /zeng2-6/ncdat/127/na/na-NA-00000001  
5: /zeng2-6/ncdat/127/na/na-NA-00000001  
6: /zeng2-6/ncdat/127/na/na-NA-00000001

Print: 100 is the number of results produced by chance to have a  
score greater than or equal to the score of the result being printed.  
and is derived by analysis of the total score distribution.

SUMMARY

Result	Seq	Score	Match	Length	Print	Description
1	1	1.6	1447	1	US-09-016-434-1493	Sequence 2, Appl
2	2	1.6	1447	1	US-09-016-434-1493	Sequence 2, Appl
3	3	1.6	1447	1	US-09-016-434-1493	Sequence 2, Appl
4	4	1.6	1447	1	US-09-016-434-1493	Sequence 2, Appl
5	5	1.6	1447	1	US-09-016-434-1493	Sequence 2, Appl
6	6	1.6	1447	1	US-09-016-434-1493	Sequence 2, Appl
7	7	1.6	1447	1	US-09-016-434-1493	Sequence 2, Appl
8	8	1.6	1447	1	US-09-016-434-1493	Sequence 2, Appl
9	9	1.6	1447	1	US-09-016-434-1493	Sequence 2, Appl
10	10	1.6	1447	1	US-09-016-434-1493	Sequence 2, Appl
11	11	1.6	1447	1	US-09-016-434-1493	Sequence 2, Appl
12	12	1.6	1447	1	US-09-016-434-1493	Sequence 2, Appl
13	13	1.6	1447	1	US-09-016-434-1493	Sequence 2, Appl
14	14	1.6	1447	1	US-09-016-434-1493	Sequence 2, Appl
15	15	1.6	1447	1	US-09-016-434-1493	Sequence 2, Appl
16	16	1.6	1447	1	US-09-016-434-1493	Sequence 2, Appl
17	17	1.6	1447	1	US-09-016-434-1493	Sequence 2, Appl
18	18	1.6	1447	1	US-09-016-434-1493	Sequence 2, Appl
19	19	1.6	1447	1	US-09-016-434-1493	Sequence 2, Appl
20	20	1.6	1447	1	US-09-016-434-1493	Sequence 2, Appl
21	21	1.6	1447	1	US-09-016-434-1493	Sequence 2, Appl
22	22	1.6	1447	1	US-09-016-434-1493	Sequence 2, Appl
23	23	1.6	1447	1	US-09-016-434-1493	Sequence 2, Appl
24	24	1.6	1447	1	US-09-016-434-1493	Sequence 2, Appl
25	25	1.6	1447	1	US-09-016-434-1493	Sequence 2, Appl
26	26	1.6	1447	1	US-09-016-434-1493	Sequence 2, Appl
27	27	1.6	1447	1	US-09-016-434-1493	Sequence 2, Appl



Best local similarity 100%, Prod. No. 4.37  
Matches 100% Conservative 0% Mismatches 0% Indels 0% Gaps 0%

QY 692 TAACTGTTAATTTCTGAG 919

DB 1543 TAACTGTTAATTTCTGAG 1562

# RESULT 4

US-08-973-297-2  
Sequence 2: Application US/08/973-297-2  
Patent No. 5626542  
GENERAL INFORMATION:  
APPLICANT: Smith, Daniel S.  
APPLICANT: Walker, John C.  
TITLE OF INVENTION: Glycine and Phascolus  
TITLE OF INVENTION: alpha D-Galactosidases  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Raising, Blending, Patrick & Perry  
CITY: Troy  
STATE: Michigan  
COUNTRY: US  
ZIP: 48060-4132  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC DOS/MS DOS  
SOFTWARE: Patent in Release #1.0, Version #1.10  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US-08-973-297-2  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Kohn, Kenneth  
REGISTRATION NUMBER: 30,955  
REFERENCE/DOCKET NUMBER: 0394,0050  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (810) 539-5050  
TELEFAX: (810) 539-5055  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1745 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

Query Match 100% Score 257 DB 37 Length 1945  
Best local similarity 100%, Prod. No. 4.37  
Matches 100% Conservative 0% Mismatches 0% Indels 0% Gaps 0%

QY 692 TAACTGTTAATTTCTGAG 919

DB 1543 TAACTGTTAATTTCTGAG 1562

# RESULT 5

US-08-973-297-2  
Sequence 2: Application US/08/973-297-2  
Patent No. 5626542  
GENERAL INFORMATION:  
APPLICANT: Smith, Daniel S.  
APPLICANT: Walker, John C.  
TITLE OF INVENTION: Glycine and Phascolus  
TITLE OF INVENTION: alpha D-Galactosidases  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kohn & Associates  
CITY: Farmington Hills  
STATE: Michigan  
COUNTRY: US  
ZIP: 48334  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC DOS/MS DOS  
SOFTWARE: Patent in Release #1.0, Version #1.10  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/973-297-2  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Kohn, Kenneth  
REGISTRATION NUMBER: 30,955  
REFERENCE/DOCKET NUMBER: 0394,0050  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (810) 539-5050  
TELEFAX: (810) 539-5055  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1745 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

Query Match 100% Score 257 DB 37 Length 1945  
Best local similarity 100%, Prod. No. 4.37  
Matches 100% Conservative 0% Mismatches 0% Indels 0% Gaps 0%

QY 692 TAACTGTTAATTTCTGAG 919

DB 1543 TAACTGTTAATTTCTGAG 1562

# RESULT 6

US-08-973-297-2  
Sequence 2: Application US/08/973-297-2  
Patent No. 5626542  
GENERAL INFORMATION:  
APPLICANT: Smith, Daniel S.  
APPLICANT: Walker, John C.  
TITLE OF INVENTION: Glycine and Phascolus  
TITLE OF INVENTION: alpha D-Galactosidases  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kohn & Associates  
CITY: Farmington Hills  
STATE: Michigan  
COUNTRY: US  
ZIP: 48334  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC DOS/MS DOS  
SOFTWARE: Patent in Release #1.0, Version #1.10  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/973-297-2  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Kohn, Kenneth  
REGISTRATION NUMBER: 30,955  
REFERENCE/DOCKET NUMBER: 0394,0050  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (810) 539-5050  
TELEFAX: (810) 539-5055  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1745 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

COUNTRY: US  
ZIP: 48334  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC DOS/MS DOS  
SOFTWARE: Patent in Release #1.0, Version #1.10  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/973-297-2  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Kohn, Kenneth  
REGISTRATION NUMBER: 30,955  
REFERENCE/DOCKET NUMBER: 0394,0050  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (810) 539-5050  
TELEFAX: (810) 539-5055  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1745 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-973-297-2

Query Match 100% Score 257 DB 37 Length 1945  
Best local similarity 100%, Prod. No. 4.37  
Matches 100% Conservative 0% Mismatches 0% Indels 0% Gaps 0%

QY 692 TAACTGTTAATTTCTGAG 919

DB 1543 TAACTGTTAATTTCTGAG 1562

# RESULT 6

US-08-973-297-2  
Sequence 2: Application US/08/973-297-2  
Patent No. 5626542  
GENERAL INFORMATION:  
APPLICANT: Smith, Daniel S.  
APPLICANT: Walker, John C.  
TITLE OF INVENTION: Glycine and Phascolus  
TITLE OF INVENTION: alpha D-Galactosidases  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kohn & Associates  
CITY: Farmington Hills  
STATE: Michigan  
COUNTRY: US  
ZIP: 48334  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC DOS/MS DOS  
SOFTWARE: Patent in Release #1.0, Version #1.10  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/973-297-2  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Kohn, Kenneth  
REGISTRATION NUMBER: 30,955  
REFERENCE/DOCKET NUMBER: 0394,0050  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (810) 539-5050  
TELEFAX: (810) 539-5055  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1745 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

PCT-US96 06511 2

Query Match 1.61, Score 20, DB 5, Length 1745  
 Best Local Similarity 100.0%, Pred. No. 4, 17  
 Matches 20, Conservative 0, Mismatches 0, Indels 0

QY 648 TATGTTTATTTTGGG 717  
 ||||| ||||| |||||  
 DB 1043 TATGTTTATTTTGGG 1562

RESULT 7  
 US-09-620 412D 65876  
 Sequence 618, Application US/59620312D  
 Patent No. 658462

GENERAL INFORMATION:

APPLICANT: Tang, Y. Tom  
 APPLICANT: Liu, Chonghua  
 APPLICANT: Asanaka, Yuhiko  
 APPLICANT: Zhang, Jie  
 APPLICANT: Ren, Fajun  
 APPLICANT: Chen, Baicheng  
 APPLICANT: Zhao, Qing A.  
 APPLICANT: Ahrman, Tom  
 APPLICANT: Xue, Aizong J.  
 APPLICANT: Yang, Yanguang  
 APPLICANT: Wang, Jiansu  
 APPLICANT: Zhou, Ping  
 APPLICANT: Wu, Yuchang  
 APPLICANT: Wang, Junxi  
 APPLICANT: Wang, Zhilwei  
 APPLICANT: Chen, Tingliang  
 APPLICANT: Primak, Radoje T.  
 TITLE OF INVENTION: NO. 650662e1 Nucleic Acids and  
 TITLE OF INVENTION: Polypeptides  
 FILE REFERENCE: 78401P28  
 CURRENT APPLICATION NUMBER: US/09/620,312D  
 CURRENT FILING DATE: 2000-07-19  
 PRIOR APPLICATION NUMBER: 09/592,317  
 PRIOR FILING DATE: 2000-04-25  
 PRIOR APPLICATION NUMBER: 09/468,725  
 PRIOR FILING DATE: 2000-01-21  
 NUMBER OF SEQ. NO. 1105  
 SOFTWARE: PCT-Genes Version 1.0  
 SEQ. NO. 658  
 LENGTH: 948  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURES:  
 NAME/KEY: CDS  
 LOCATION: 1941..12064

US-09-620 412D 65876

Query Match 1.61, Score 20, DB 4, Length 1745  
 Best Local Similarity 100.0%, Pred. No. 4, 17  
 Matches 20, Conservative 0, Mismatches 0, Indels 0

QY 671 TTTTATAGCTGTTATTT 711  
 ||||| ||||| |||||  
 DL 2413 TTTTATAGCTGTTATTT 2194

RESULT 8  
 US-09-643 497150  
 Sequence 150, Application US/09643597  
 Patent No. 645372

GENERAL INFORMATION:

APPLICANT: Wang, Tongtong  
 APPLICANT: Ren, Liqun  
 APPLICANT: Kules, Michael D.  
 APPLICANT: Bunker, Chaitanya S.  
 APPLICANT: Hosken, Nancy  
 APPLICANT: Fanger, Gary R.

APPLICANT: Tan, Samuel Z.  
 APPLICANT: Wang, Aimin  
 APPLICANT: Sakhya, Vignu A.M.  
 APPLICANT: Wamberson, Daniel A.  
 APPLICANT: Morrison, Patricia C.  
 TITLE OF INVENTION: COMBINATIONS AND METHODS FOR THE THERAPY  
 TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER  
 FILE REFERENCE: 21011465C1  
 CURRENT APPLICATION NUMBER: US/09/643,597  
 CURRENT FILING DATE: 2000-08-22  
 NUMBER OF SEQ. NO. 21  
 SOFTWARE: PCT-Genes Version 1.0  
 SEQ. NO. 150  
 LENGTH: 301  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURES:  
 NAME/KEY: CDS  
 LOCATION: 1941..12064

Query Match 1.61, Score 20, DB 4, Length 1745  
 Best Local Similarity 100.0%, Pred. No. 4, 17  
 Matches 20, Conservative 0, Mismatches 0, Indels 0

QY 476 AGGAGATCTTCTTGA 100  
 ||||| ||||| |||||  
 DB 346 AGGAGATCTTCTTGA 484

RESULT 9  
 US-09-460 494A 113  
 Sequence 113, Application US/09460494A  
 Patent No. 4948191

GENERAL INFORMATION:  
 APPLICANT: Wang, Dandan  
 APPLICANT: Fan, Liqun  
 APPLICANT: Hosken, Nancy A.  
 APPLICANT: Filler, Yehouda D.  
 APPLICANT: Fanger, Gary R.  
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY  
 TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER  
 FILE REFERENCE: 21011465C1  
 CURRENT APPLICATION NUMBER: US/09/460,494A  
 CURRENT FILING DATE: 2000-04-27  
 NUMBER OF SEQ. NO. 113  
 SOFTWARE: PCT-Genes Version 1.0  
 SEQ. NO. 113  
 LENGTH: 191  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURES:  
 NAME/KEY: CDS  
 LOCATION: 1941..12064

Query Match 1.61, Score 19, DB 4, Length 1745  
 Best Local Similarity 100.0%, Pred. No. 4, 17  
 Matches 19, Conservative 0, Mismatches 1, Indels 0

QY 476 AGGAGATCTTCTTGA 100  
 ||||| ||||| |||||  
 DB 346 AGGAGATCTTCTTGA 484

RESULT 10  
 US-09-643 497150  
 Sequence 150, Application US/09643597  
 Patent No. 645372

GENERAL INFORMATION:  
 APPLICANT: Wang, Tongtong  
 APPLICANT: Ren, Liqun  
 APPLICANT: Kules, Michael D.  
 APPLICANT: Bunker, Chaitanya S.  
 APPLICANT: Hosken, Nancy A.  
 APPLICANT: Fanger, Gary R.  
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY  
 TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER  
 FILE REFERENCE: 21011465C1



1 CURRENT APPLICATION NUMBER: US/09/542,615A  
 2 CURRENT FILING DATE: 2001-04-14  
 3 NUMBER OF SEQ ID NOS: 358  
 4 SOFTWARE: FASTSEQ for Windows Version 4.0  
 5 SEQ ID NO 150  
 6 LENGTH: 171  
 7 TYPE: DNA  
 8 ORGANISM: Homo sapiens  
 9 US 09/542,615A 150

Query Match 150: Score: 150; Length: 171  
 Best Local Similarity: 100.0%; Ident. No. 13;  
 Matches: 19; Conservation: 0; Mismatches: 0; Gaps: 0;

QY 346 ACAGATATTTT TTA 150  
 DB 346 ACAGATATTTT TTA 150

RESULT 1:  
 US 09/542,615A 150  
 1 Sequence: 150, Application US/09/542,615A  
 2 Patent No. 6420135

#### GENERAL INFORMATION:

1 APPLICANT: Wang, Tonglong  
 2 APPLICANT: Fan, Liqun  
 3 APPLICANT: Kalos, Michael D.  
 4 APPLICANT: Rantur, Chaitanya S.  
 5 APPLICANT: Rosen, Nancy  
 6 APPLICANT: Farney, Gary A.  
 7 APPLICANT: Li, Samuel X.  
 8 APPLICANT: Wang, Aijun  
 9 APPLICANT: Skeiky, Yael A.W.  
 10 TITLE OF INVENTION: COMPOSITION AND METHODS FOR THE THERAPY  
 11 TITLE OF INVENTION: AND DIAGNOSIS OF DNA DAMAGE  
 12 FILE REFERENCE: EP01 43479  
 13 CURRENT APPLICATION NUMBER: US/09/ 4,415  
 14 NUMBER OF SEQ ID NOS: 358  
 15 SOFTWARE: FASTSEQ for Windows Version 4.0  
 16 SEQ ID NO 150  
 17 LENGTH: 171  
 18 TYPE: DNA  
 19 ORGANISM: Homo sapiens  
 20 US 09/542,615A 150

Query Match 150: Score: 150; Length: 171  
 Best Local Similarity: 100.0%; Ident. No. 13;  
 Matches: 19; Conservation: 0; Mismatches: 0; Gaps: 0;

QY 346 ACAGATATTTT TTA 150  
 DB 346 ACAGATATTTT TTA 150

RESULT 1:  
 US 09/542,615A 150  
 1 Sequence: 150, Application US/09/542,615A  
 2 Patent No. 6420135

#### GENERAL INFORMATION:

1 APPLICANT: Charles Kusch  
 2 TITLE OF INVENTION: SUPEROXIDE DISMUTASE POLYMERIZABLES AND SEQUENCES  
 3 NUMBER OF SEQUENCES: 351  
 4 CORRESPONDENCE ADDRESS:  
 5 ADDRESSEE: Human Genome Sciences, Inc.  
 6 STREET: 9410 Key West Avenue  
 7 CITY: Rockville  
 8 STATE: Maryland  
 9 COUNTRY: USA  
 10 ZIP: 20850  
 11 COMPUTER READABLE FORM:  
 12 MEDIUM TYPE: Diskette, 3.50 inch, 1.44M storage  
 13 COMPUTER: HP Vectra 486/33

1 OPERATING SYSTEM: MACOS VERSION 6.2  
 2 SOFTWARE: ASCII Text  
 3 CURRENT APPLICATION DATA:  
 4 APPLICATION NUMBER: US/09/542,615A  
 5 FILING DATE:  
 6 CLASSIFICATION: 424  
 7 PRIOR APPLICATION DATA:  
 8 APPLICATION NUMBER:  
 9 FILING DATE:  
 10 ATTORNEY/AGENT INFORMATION:  
 11 NAME: Brooks A. Anders  
 12 REGISTRATION NUMBER: 36,150  
 13 REFERENCE/DOCKET NUMBER: BR-4461  
 14 TELECOMMUNICATION INFORMATION:  
 15 TELEPHONE: (301) 309-9544  
 16 TELEFAX: (301) 309-9512  
 17 INFORMATION FOR SEQ ID NO: 150  
 18 SEQUENCE CHARACTERISTICS:  
 19 LENGTH: 1659; base pairs  
 20 TYPE: nucleic acid  
 21 STRANDEDNESS: double  
 22 TOPOLOGY: linear  
 23 US-09/542,615A 150

Query Match 150: Score: 150; Length: 1659  
 Best Local Similarity: 100.0%; Ident. No. 13;  
 Matches: 19; Conservation: 0; Mismatches: 0; Gaps: 0;

QY 11684 TCTCAGATGAGGATTTT 1659  
 DB 11684 TCTCAGATGAGGATTTT 1659

RESULT 1:  
 US 08/438,540-170  
 1 Sequence: 1, Application US/08/438,540-170  
 2 Patent No. 5470318

#### GENERAL INFORMATION:

1 APPLICANT: Herri Tledge and Jurgen Prosius  
 2 TITLE OF INVENTION: BCDORNA, Probes Therefor and the Use Thereof  
 3 NUMBER OF SEQUENCES: 13  
 4 CORRESPONDENCE ADDRESS:  
 5 ADDRESSEE: Rumbaugh, Graves, Donahue & Raymond  
 6 STREET: 10 Rockefeller Plaza  
 7 CITY: New York  
 8 STATE: New York  
 9 COUNTRY: USA  
 10 ZIP: 10012  
 11 COMPUTER READABLE FORM:  
 12 MEDIUM TYPE: 3.5 inch floppy diskette  
 13 COMPUTER: IBM PC COMPATIBLE  
 14 OPERATING SYSTEM: DOS VER. 5.0  
 15 SOFTWARE: WP 5.1  
 16 CURRENT APPLICATION DATA:  
 17 APPLICATION NUMBER: US/08/438,540-170  
 18 FILING DATE:  
 19 PRIOR APPLICATION DATA:  
 20 APPLICATION NUMBER: US/08/438,540-170  
 21 FILING DATE:  
 22 ATTORNEY/AGENT INFORMATION:  
 23 NAME: Martina T. Larson  
 24 REGISTRATION NUMBER: 32,038  
 25 REFERENCE/DOCKET NUMBER: 29407  
 26 TELECOMMUNICATION INFORMATION:  
 27 TELEPHONE: (212) 406-2500  
 28 TELEFAX: (212) 765-2519  
 29 INFORMATION FOR SEQ ID NO: 13  
 30 SEQUENCE CHARACTERISTICS:  
 31 LENGTH: 200  
 32 TYPE: nucleic acid  
 33 STRANDEDNESS: single  
 34 TOPOLOGY: linear

Query Match 13: Score: 13; Length: 200  
 Best Local Similarity: 100.0%; Ident. No. 13;  
 Matches: 19; Conservation: 0; Mismatches: 0; Gaps: 0;

QY 11684 TCTCAGATGAGGATTTT 1659  
 DB 11684 TCTCAGATGAGGATTTT 1659

RESULT 1:  
 US 08/438,540-170  
 1 Sequence: 1, Application US/08/438,540-170  
 2 Patent No. 5470318

#### GENERAL INFORMATION:

1 APPLICANT: Herri Tledge and Jurgen Prosius  
 2 TITLE OF INVENTION: BCDORNA, Probes Therefor and the Use Thereof  
 3 NUMBER OF SEQUENCES: 13  
 4 CORRESPONDENCE ADDRESS:  
 5 ADDRESSEE: Rumbaugh, Graves, Donahue & Raymond  
 6 STREET: 10 Rockefeller Plaza  
 7 CITY: New York  
 8 STATE: New York  
 9 COUNTRY: USA  
 10 ZIP: 10012  
 11 COMPUTER READABLE FORM:  
 12 MEDIUM TYPE: 3.5 inch floppy diskette  
 13 COMPUTER: IBM PC COMPATIBLE  
 14 OPERATING SYSTEM: DOS VER. 5.0  
 15 SOFTWARE: WP 5.1  
 16 CURRENT APPLICATION DATA:  
 17 APPLICATION NUMBER: US/08/438,540-170  
 18 FILING DATE:  
 19 PRIOR APPLICATION DATA:  
 20 APPLICATION NUMBER: US/08/438,540-170  
 21 FILING DATE:  
 22 ATTORNEY/AGENT INFORMATION:  
 23 NAME: Martina T. Larson  
 24 REGISTRATION NUMBER: 32,038  
 25 REFERENCE/DOCKET NUMBER: 29407  
 26 TELECOMMUNICATION INFORMATION:  
 27 TELEPHONE: (212) 406-2500  
 28 TELEFAX: (212) 765-2519  
 29 INFORMATION FOR SEQ ID NO: 13  
 30 SEQUENCE CHARACTERISTICS:  
 31 LENGTH: 200  
 32 TYPE: nucleic acid  
 33 STRANDEDNESS: single  
 34 TOPOLOGY: linear

```

1 MOLECULE TYPE: MRNA
2 HYPOTHETICAL: no
3 ANTI-SENSE: no
4 ORIGINAL SOURCE:
5 ORGANISM: human
6 STRAIN:
7 INDIVIDUAL ISOLATE:
8 DEVELOPMENTAL STAGE:
9 HAPLOTYPE:
10 TISSUE TYPE: neocortex
11 CELL TYPE:
12 CELL LINE:
13 ORGANELLE:
14 FEATURE:
15 NAME/KEY: Human BC200 RNA
16 US CB 438-4381
17
18 Query Match: 1.43; Score 18; BB 1; Length 200.
19 Best Local Similarity: 100%; Pct No. 40.
20 Matches 18; Conservative 0; Mismatches 0; Gaps 0.
21
22 QY 100 PATENTTTTTTGGCTTC 143
23 DE 100 PATATTTTGTGCTTC 143
24
25 RESULT 14
26 US CB 438-4381
27 Sequence 1: Application US/0847442
28 Patent No. 5366128
29 GENERAL INFORMATION:
30 APPLICANT: Tiedje, Henry
31 TITLE OF INVENTION: BCCORNA, Probes Therefor and the
32 Title of Invention: Use thereof
33 NUMBER OF REFERENCES: 13
34 CORRESPONDENCE ADDRESS:
35 ADDRESSEE: Blumberg, Graves, Donohue & Raymond
36 STREET: 30 Rockefeller Plaza
37 CITY: New York
38 STATE: New York
39 COUNTRY: USA
40 ZIP: 10017
41 COMPUTER READABLE FORM:
42 MEDIAN TYPE: 4AMS 3-1/2 inch diskette
43 COMPUTER, IBM OR COMPATIBLE
44 OPERATING SYSTEM: DOS VER. 5.0
45 SOFTWARE: ASCII
46 CURRENT APPLICATION DATA:
47 APPLICATION NUMBER: US/08477442
48 FILING DATE: 27-JUNE-1995
49 PRIOR APPLICATION DATA:
50 APPLICATION NUMBER:
51 FILING DATE:
52 ATTORNEY/AGENT INFORMATION:
53 NAME: Lisa R Kolt
54 REGISTRATION NUMBER: 35,245
55 REFERENCE/SOCKET NUMBER: 29407 PWO 1154/1143
56 TELECOMMUNICATION INFORMATION:
57 TELEPHONE: (212) 406-2500
58 TELEFAX: (212) 365-2519
59 TELEX: 246-55
60 INFORMATION FOR SEQ ID NO: 1:
61 SEQUENCE CHARACTERISTICS:
62 LENGTH: 200
63 TYPE: nucleic acid
64 STRANDEDNESS: single
65 TOPOLOGY: linear
66 MOLECULE TYPE: MRNA
67 HYPOTHETICAL: no
68 ANTI-SENSE: no
69 ORIGINAL SOURCE:
70 ORGANISM: human

```

1 CHILL LINE  
2 GANGLER  
3 FEATURES  
4 NAMEKEY: human BC201 FSA  
5 OCT 1994 09:12:11

Query Match: 1.43; Score 14; DB 4; Length 602;  
Best Local Similarity: 100.0%; Pred. No. 40;  
Matches: 18; Conservative: 0; Mismatches: 0; Indels: 0; Gaps: 0;

QY 14 TTTACTTTTCTTTTCA 124  
DB 14 TTTACTTTTCTTTTCA 124

RESULT 19

US 09-018 294-27/C  
1 Sequence 17, Application US/2904234  
2 Patent No. 6265211  
3 GENERAL INFORMATION:

4 APPLICANT: Choo, Keng Heng Andy  
5 APPLICANT: Choo, Keng Heng Andy  
6 APPLICANT: Choo, Keng Heng Andy  
7 TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE  
8 FILE REPRESENTATIVE: Davies Co.  
9 CURRENT APPLICATION NUMBER: US/09/0178,194  
10 PRIOR FILING DATE: 1994-09-14  
11 NUMBER OF SEQ ID NOS: 29  
12 SOFTWARE: PatentLit Ver. 3.0  
13 SEQ ID NO 27  
14 LENGTH: 602  
15 TYPE: DNA  
16 ORGANISM: Bacillus subtilis 47 (fragment)  
17 US 09-018 294-27

Query Match: 1.43; Score 14; DB 4; Length 602;  
Best Local Similarity: 100.0%; Pred. No. 40;  
Matches: 18; Conservative: 0; Mismatches: 0; Indels: 0; Gaps: 0;

QY 14 TTTACTTTTCTTTTCA 124  
DB 14 TTTACTTTTCTTTTCA 124

RESULT 19

US 09-018 294-27/C  
1 Sequence 17, Application US/2904234  
2 Patent No. 6265211  
3 GENERAL INFORMATION:

4 APPLICANT: Choo, Keng Heng Andy  
5 APPLICANT: Choo, Keng Heng Andy  
6 APPLICANT: Choo, Keng Heng Andy  
7 TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE  
8 FILE REPRESENTATIVE: Davies Co.  
9 CURRENT APPLICATION NUMBER: US/09/0178,194  
10 PRIOR FILING DATE: 1994-09-14  
11 NUMBER OF SEQ ID NOS: 29  
12 SOFTWARE: PatentLit Ver. 3.0  
13 SEQ ID NO 27  
14 LENGTH: 602  
15 TYPE: DNA  
16 ORGANISM: Bacillus subtilis 47 (fragment)  
17 US 09-018 294-27

Query Match: 1.43; Score 14; DB 4; Length 602;  
Best Local Similarity: 100.0%; Pred. No. 40;  
Matches: 18; Conservative: 0; Mismatches: 0; Indels: 0; Gaps: 0;

QY 43 TTTACTTTTCTTTTCA 124  
DB 43 TTTACTTTTCTTTTCA 124

RESULT 19

US 09-018 294-27/C  
1 Sequence 17, Application US/2904234  
2 Patent No. 6265211  
3 GENERAL INFORMATION:

4 APPLICANT: Choo, Keng Heng Andy  
5 APPLICANT: Choo, Keng Heng Andy  
6 APPLICANT: Choo, Keng Heng Andy  
7 TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE  
8 FILE REPRESENTATIVE: Davies Co.  
9 CURRENT APPLICATION NUMBER: US/09/0178,194  
10 PRIOR FILING DATE: 1994-09-14  
11 NUMBER OF SEQ ID NOS: 29  
12 SOFTWARE: PatentLit Ver. 3.0  
13 SEQ ID NO 27  
14 LENGTH: 602  
15 TYPE: DNA  
16 ORGANISM: Bacillus subtilis 47 (fragment)  
17 US 09-018 294-27

Query Match: 1.43; Score 18; DB 3; Length 602;  
Best Local Similarity: 100.0%; Pred. No. 40;  
Matches: 18; Conservative: 0; Mismatches: 0; Indels: 0; Gaps: 0;

QY 119 TTTACTTTTCTTTTCA 124  
DB 43 TTTACTTTTCTTTTCA 413

RESULT 19

US 09-018 294-27/C  
1 Sequence 17, Application US/2904234  
2 Patent No. 6265211  
3 GENERAL INFORMATION:

4 APPLICANT: Choo, Keng Heng Andy  
5 APPLICANT: Choo, Keng Heng Andy  
6 APPLICANT: Choo, Keng Heng Andy  
7 TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE  
8 FILE REPRESENTATIVE: Davies Co.  
9 CURRENT APPLICATION NUMBER: US/09/0178,194  
10 PRIOR FILING DATE: 1994-09-14  
11 NUMBER OF SEQ ID NOS: 29  
12 SOFTWARE: PatentLit Ver. 3.0  
13 SEQ ID NO 27  
14 LENGTH: 602  
15 TYPE: DNA  
16 ORGANISM: Bacillus subtilis 47 (fragment)  
17 US 09-018 294-27

Query Match: 1.43; Score 18; DB 3; Length 602;  
Best Local Similarity: 100.0%; Pred. No. 40;  
Matches: 18; Conservative: 0; Mismatches: 0; Indels: 0; Gaps: 0;

QY 119 TTTACTTTTCTTTTCA 124  
DB 43 TTTACTTTTCTTTTCA 413

## RESULT 20

US-08-976-269-9

Sequence 3: Application US/08976269

Patent No. 6116609

GENERAL INFORMATION:

APPLICANT: Ellison, Patrick J

APPLICANT: Gilio, Gil H.

APPLICANT: Welch, Rodney A.

TITLE OF INVENTION: Nucleotide Sequence of Escherichia coli

Patent No. 6116609

NUMBER OF SEQUENCES: 142

CORRESPONDENCE ADDRESS:

ADDRESSEE: Streiner, Kessler, Goldstein &amp; Fox P.L.L.C.

STREET: 1100 New York Ave, N.W., Suite 400

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20005-1414

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1 4MB at 1100

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 5.0

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/976,269

FILING DATE: Hologith

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/031,616 AND US 60/061,954

ATTORNEY/AGENT INFORMATION:

NAME: Swift, Eric K.

REGISTRATION NUMBER: 36,669

REFERENCE/DOCKET NUMBER: 1488-0740-02/PSEUDOM

TELECOMMUNICATION INFORMATION:

TELEPHONE: 12021 371-2600

TELEFAX: 12021 371-2640

INFORMATION FOR SEQ ID NO: 9

SEQUENCE CHARACTERISTICS:

LENGTH: 120 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPLOGY: Linear

US-08-976-269-9

## Query Match

Best Local Similarity 100.0%; Seed No. 46; Length 720;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 242 GTGATCCAGAGTTCTT 264

|||||

Dd 141 GTGATCCAGAGTTCTT 138

## RESULT 21

US-09-252-991A-12131

Sequence 10131: Application US/09252991A

Patent No. 6461795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 101196 136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-03-18

PRIOR APPLICATION NUMBER: US 60/074,784

PRIOR FILING DATE: 1998-02-19

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 12131:

LENGTH: 1413

TYPE: JNA

ORGANISM: Pseudomonas aeruginosa

## US-09-252-991A-12131

## Query Match

Best Local Similarity 100.0%; Seed No. 46; Length 1413;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 423 AACAGATGATGATG 440

|||||

Dd 176 AACAGATGATGATG 191

## RESULT 22

US-09-252-991A-12131

Sequence 12131: Application US/09252991A

Patent No. 6461795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 101196 136

CURRENT APPLICATION NUMBER: US/09/252,991A

PRIOR APPLICATION NUMBER: US 60/074,784

PRIOR FILING DATE: 1998-02-19

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 12131:

LENGTH: 1413

TYPE: JNA

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-12131

## Query Match

Best Local Similarity 100.0%; Seed No. 46; Length 1493;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 413 AAAAGTTGATG 443

|||||

Dd 1156 AAAAGTTGATG 131

## RESULT 23

US-09-403-6286-1-7

Sequence 7: Application US/094036286

Patent No. 6424154

GENERAL INFORMATION:

APPLICANT: BRIDGES, BARRY E.

TITLE OF INVENTION: CYRIL FULLMANNER

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESS: BRIDGES PART OF NEWPORT AND NEWRY

STREET: 107 VANDER STREET

CITY: NEWPORT

STATE: NEWCASTLE

COUNTRY: U.K.

DATE: 1994

COMPLETES DEPOSABLE FORM

TITLE OF INVENTION: CYRIL FULLMANNER

OPERATING SYSTEM: MICROSOFT WINDOWS 95

SOFTWARE: MICROSOFT WORD 6.0 WINDO 95 (1.0)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/403,628A

FILING DATE: 1998-03-18

CLASSIFICATION: C08K000000

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/403,628A

FILING DATE: MAY 6, 1997

ATTORNEY/AGENT INFORMATION:

NAME: MALABIAN, WILLIAM F.

REGISTRATION NUMBER: 41173

REFERENCE/DOCKET NUMBER: BP 1158

```

1 PRIOR FILING DATE: 1995 05 17
2 PRIOR APPLICATION NUMBER: US 09/198,568
3 PRIOR FILING DATE: 1994 12 18
4 NUMBER OF SEQ ID NOS: 12
5 SOFTWARE: PATGEN version 3.0
6 SEQ ID NO 1
7 LENGTH: 4122
8 TYPE: DNA
9 ORGANISM: WAP gene p: mouse; fragment from the C57B/6 mouse strain
10 FEATURES:
11 NAME/KEY: misc feature
12 LOCATION: 1543..1544
13 OTHER INFORMATION: this is unknown
14 NAME/KEY: misc feature
15 LOCATION: 1547..1549
16 OTHER INFORMATION: this is unknown
17 US 99-321541

```

```

Query Match      : 94% Score 187 CR 10 Length 41227
Best Local Similarity : 100% Pctd No. 40
Matches 15 Conservative 0 Mismatches 0 Indels 0 Gaps 0

```

100 TAAGTTCATATTATGTT 1184  
|||||  
100 TAAGTTCATATTATGTT 60

RESULT 26  
US-05-078-194-7/C  
Sequence V. Application US/09/078214  
Patent No. 6265211  
GENERAL INFORMATION:  
APPLICANT: CHOO, KONG-HONG Andy  
APPLICANT: DO SALT, Denise  
APPLICANT: CARRILLO, MICHAEL R.  
TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE  
FILE REFERENCE: DAVIES CO.  
CURRENT APPLICATION NUMBER: 06/04/2008, 154  
CURRENT FILING DATE: 1998 01-13  
NUMBER OF SEQ ID NOS: 19

```

? SEQ ID NO : 7 accession: J01698
? LENGTH: 11411
? TYPE: CNA
? ORGANISM: BAC-PE COMING 3
US-09 078 2947
      Query Match       144% Score 16. DE 1; Length 10113;
Best Local Similarity Found 144% Dec No. 40;
Matched IR Conservation Q; Miss-matches O; Indels M; Gaps G
          QQ    1199 TTTTCTTTTTTTTTTG   1216
              ||| | | | | | | | |
             1122 TTACTTTTTTTGTA   1159

RESULT 77
US-09-749-3BA+3
Sequence 3 Application US/97495B8
Parent No. 6423521
GENERAL INFORMATION:
APPLICANT: CHANDRANAGULISWARAN, Ishwar et al
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
FILE REFERENCE: THEREOF
FILE REFERENCE: CLC01068
CURRENT APPLICATION NUMBER: US/05/749,588
CURRENT FILING DATE: 2000.12.28
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 3
LENGTH: 36159
TYPE: DNA
```





## TITLE OF INVENTION: METHOD FOR IDENTIFYING METASTATIC SEQUENCES

NUMBER OF SEQUENCES: 175  
 CORRESPONDENCE ADDRESS:  
 ADDRESSER: BAKER & BOTTIS, L.L.P.  
 STREET: 1299 Pennsylvania Avenue, N.W.  
 CITY: Washington  
 STATE: DC  
 COUNTRY: USA  
 ZIP: 20004-2400

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FASTSEQ Version 1.5  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/05/594,031  
 FILING DATE: 30 JAN-1998  
 CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 607201,514  
 FILING DATE: 16 NOV-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Seward, K. Jones  
 REGISTRATION NUMBER: 44,162  
 REFERENCE/AGENT NUMBER: 01446,0112  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 202-639-7700  
 TELEFAX: 202-639-7690  
 TELEX:

INFORMATION FOR SEQ ID NO: 154

SEQUENCE CHARACTERISTICS:  
 LENGTH: 434 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULAR TYPE: cDNA  
 HYDROPHOBIC: NO  
 ANTI-SENSE: NO  
 PROMOTER TYPE:  
 ORIGINAL SOURCE:  
 US 08-594-031, 301

Query Match: 1.00, Score 171, ID 11, Length 404,  
 Best Local Similarity: 100.0%, Pctd. No. 1,464,021,  
 Matches: 171, Conservative: 97, Mismatches: 0, Indels: 0

QY 1009 TTGTAAATACCTCAGCA 1016  
 111 111 111  
 DB 1006 TTGTAAATACCTCAGCA 172

RESULT 33  
 US 09-309,317 1  
 Sequence 1, Application US/0930317  
 Patent No. 627197  
 GENERAL INFORMATION:  
 APPLICANT: Plusner, Stanley  
 APPLICANT: Tenbray, Patrick  
 APPLICANT: Moore, Richard  
 APPLICANT: Westaway, David  
 APPLICANT: Hoon, Leroy E.  
 APPLICANT: Lee, Inyong  
 TITLE OF INVENTION: Pip-like Gene  
 FILE REFERENCE: 6510-330031  
 CURRENT APPLICATION NUMBER: US/09/309,317  
 CURRENT FILING DATE: 1999-05-11  
 NUMBER OF SEQ ID NOS: 21  
 SOFTWARE: FASTSEQ for Windows Version 4.0  
 SEQ ID NO 1  
 LENGTH: 540  
 TYPE: DNA  
 ORGANISM: mus musculus  
 US-09-309-317-1

Query Match: 1.00, Score 171, ID 11, Length 540,  
 Best Local Similarity: 100.0%, Pctd. No. 1,464,021,  
 Matches: 171, Conservative: 97, Mismatches: 0, Indels: 0

QY 1006 TTGTAAATACCTCAGCA 1016  
 111 111 111  
 DB 1006 TTGTAAATACCTCAGCA 172

## RESULT 34

US 09-252-391A 1009  
 Sequence 175, Application US/09252391A  
 Patent No. 6461926  
 GENERAL INFORMATION:

APPLICANT: MARCH, J. ROBERT  
 TITLE OF INVENTION: NUCLEIC ACID AND ANALOGUE SEQUENCES RELATING TO PROTEINASES  
 TITLE OF INVENTION: ANALOGUE OF PROTEINASES AND THERAPEUTICS  
 FILE REFERENCE: 101010101

CURRENT APPLICATION NUMBER: 09/000,000  
 CURRENT FILING DATE: 10/10/00  
 PRIOR APPLICATION NUMBER: 09/000,000  
 PRIOR FILING DATE: 10/10/00  
 PRIOR APPLICATION NUMBER: 09/000,000  
 PRIOR FILING DATE: 10/10/00  
 NUMBER OF SEQ ID NOS: 100  
 SEQ ID NO 1  
 LENGTH: 600  
 TYPE: DNA  
 ORGANISM: Pseudomonas aeruginosa  
 US-09-252-391A 1009

Query Match: 1.00, Score 171, ID 11, Length 600,  
 Best Local Similarity: 100.0%, Pctd. No. 1,464,021,  
 Matches: 171, Conservative: 97, Mismatches: 0, Indels: 0

QY 1009 TTGTAAATACCTCAGCA 1016  
 111 111 111  
 DB 1006 TTGTAAATACCTCAGCA 172

RESULT 35  
 US 08-998,416 1009  
 Sequence 175, Application US/0898416  
 Patent No. 6461926  
 GENERAL INFORMATION:

APPLICANT: Plusner, Stanley  
 APPLICANT: Tenbray, Patrick  
 APPLICANT: Moore, Richard  
 APPLICANT: Westaway, David  
 APPLICANT: Hoon, Leroy E.  
 APPLICANT: Lee, Inyong  
 TITLE OF INVENTION: Pip-like Gene  
 TITLE OF INVENTION: ANALOGUE OF PROTEINASES  
 NUMBER OF SEQUENCES: 175  
 CORRESPONDENCE ADDRESS:  
 ADDRESSER: No. 6300, 101010101  
 STREET: 1004 Connecticut Road  
 CITY: Research Triangle Park  
 STATE: NC 27709  
 COUNTRY: USA  
 ZIP: 27709

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/998,416  
 FILING DATE: 24-DEC-1997  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:





QY 431 GTCGAGGACATCAACG 447  
 DB 1128 GTCGAGGACATCAACG 144

RESULT 39  
 US 09-128-452-2742  
 Sequence 2732, Application US/09124352  
 Patent No. 656,938  
 GENERAL INFORMATION:  
 APPLICANT: GARY D. Sretor et al.  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ADENYNYLTRANSFERASE  
 FILE REFERENCE: G0799 03PA  
 CURRENT APPLICATION NUMBER: US/09128-452  
 CURRENT FILING DATE: 1999-06-24  
 NUMBER OF SEQ ID NO'S: 8252  
 SEQ ID NO 2732  
 LENGTH: 1216  
 TYPE: DNA  
 ORGANISM: Adenoviridae; bacteriophage  
 US 09-128-452-2742

Query Match 1.33, Score 174, DB 41, Length 1216;  
 Best Local Similarity 100.0%; Prod. No. 100002;  
 Matches 17, Conservative 0, Mismatches 0, Indels 0, Gaps 0

QY 455 TTTCATTTTATGTS 871  
 DB 876 TTTCATTTTATGTS 844

RESULT 40  
 US-08-599-171A 172  
 Sequence 17, Application US/08599171A  
 Patent No. 591,147  
 GENERAL INFORMATION:  
 APPLICANT: WARREN, PATRICK V.  
 TITLE OF INVENTION: TRANSAMINASES AND AMINO-TRANSFERASES  
 NUMBER OF SEQUENCES: 32  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: JARELLA, BYRNE, BAIN, GIFFORD,  
 ADDRESSES: 10001, STEWART & CLUSTON  
 STREET: 6 BECKER FARM ROAD  
 CITY: ROSELAND  
 STATE: NEW JERSEY  
 COUNTRY: USA  
 ZIP: 07068

COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5 INCH DISKETTE  
 (COMPUTER IBM PS/2)  
 OPERATING SYSTEM: MS DOS  
 SOFTWARE: WORD PERFECT 5.1  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08599-171A  
 FILING DATE: Concurrently  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: HERRON, CHARLES J.  
 REGISTRATION NUMBER: 28,019  
 REFERENCE/DOCKET NUMBER: 33400-28  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 201-994-1700  
 TELEFAX: 201-994-1744  
 INFORMATION FOR SEQ ID NO: 19:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1362 NUCLEOTIDES  
 TYPE: NUCLEIC ACID

STANDARDIZED SEQUENCE  
 TOPOLONG LINEAR  
 MOLECULE TYPE: BB AND DBA  
 US 18-994-174A-17

Query Match 1.33, Score 174, DB 41, Length 1362;  
 Best Local Similarity 100.0%; Prod. No. 100002;  
 Matches 17, Conservative 0, Mismatches 0, Indels 0, Gaps 0

QY 422 AGGATGATGATTTTA 428  
 DB 457 AGGATGATGATTTTA 440

RESULT 41  
 US-08-599-171A 172  
 Sequence 17, Application US/08599171A  
 Patent No. 591,147  
 GENERAL INFORMATION:  
 APPLICANT: WARREN, PATRICK V.  
 TITLE OF INVENTION: TRANSAMINASES AND AMINO-TRANSFERASES  
 NUMBER OF SEQUENCES: 32  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: JARELLA, BYRNE, BAIN, GIFFORD,  
 STREET: 6 BECKER FARM ROAD  
 CITY: ROSELAND  
 STATE: NEW JERSEY  
 COUNTRY: USA  
 ZIP: 07068

COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5 INCH DISKETTE  
 (COMPUTER IBM PS/2)  
 OPERATING SYSTEM: MS DOS  
 SOFTWARE: WORD PERFECT 5.1  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08599-171A  
 FILING DATE: Concurrently  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: HERRON, CHARLES J.  
 REGISTRATION NUMBER: 28,019  
 REFERENCE/DOCKET NUMBER: 33400-28  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 201-994-1700  
 TELEFAX: 201-994-1744  
 INFORMATION FOR SEQ ID NO: 19:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1362 NUCLEOTIDES  
 TYPE: NUCLEIC ACID  
 MOLECULE TYPE: BB AND DBA  
 FEATURES:  
 NAME/KEY: 19113  
 LOCATION: 19113  
 US-08-599-171A

Query Match 1.33, Score 174, DB 41, Length 1362;  
 Best Local Similarity 100.0%; Prod. No. 100002;  
 Matches 17, Conservative 0, Mismatches 0, Indels 0, Gaps 0

QY 922 AGGATGATGATTTTA 938  
 DB 552 AGGATGATGATTTTA 536

RESULT 42  
US-09-069-226-19/c  
Sequence 19, Application US/09069226  
Patent No. 6013509  
GENERAL INFORMATION:  
APPLICANT: WARREN, Patrick V.  
TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
ADDRESSEE: GECCHI, STEWART & CLSTEIN  
STREET: 6 BROOKER PARK ROAD  
CITY: ROSELAND  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/069,226  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/599,172  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: HERRON, CHARLES J.  
REGISTRATION NUMBER: 28,019  
REFERENCE/DOCKET NUMBER: 33,400-38  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1744  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1362 NUCLEOTIDES  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE: GENOMIC DNA  
US-09-069-226-19

Query Match 1.3% Score 17; DB 3; Length 1362;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Cy 922 AGGAGATGGGAGTTT 936  
Db 552 AGGAGATGGGAGTTT 536  
RESULT 43  
US-09-012-184-19/c  
Sequence 19, Application US/0942184  
Patent No. 6268188  
GENERAL INFORMATION:  
APPLICANT: Warren, Patrick V.  
APPLICANT: Swanson, Ronald V.  
TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson, P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows95  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/412,184  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/646,590  
FILING DATE: 08-May-1996  
APPLICATION NUMBER: 08/599,171  
FILING DATE: 03-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/01094  
FILING DATE: 21-January 1997  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Halle, Ph.D., Lisa A.  
REGISTRATION NUMBER: 38,147  
REFERENCE/DOCKET NUMBER: 09010/017001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619/678-5070  
TELEFAX: 619/678-5099  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1362 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 1...1359  
US-09-412-184-19

Query Match 1.3% Score 17; DB 1; Length 1362;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 922 AGGAGATGGGAGTTT 938  
Db 552 AGGAGATGGGAGTTT 536

RESULT 44  
US-09-252-991A-4224/c  
Sequence 4224, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubinfeld et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMING ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,789  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/034,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 4224  
LENGTH: 1395  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-4224

Query Match 1.3% Score 17; DB 4; Length 1395;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 431 GCTGGAGGACATCAACG 447  
Db 116 GCTGGAGGACATCAACG 100

RESULT 45  
 US-09-252-991A 3651/c  
 ? Sequence 1651, Application US/09252991A  
 ? Patent No. 6551795  
 ? GENERAL INFORMATION:  
 ? APPLICANT: Marc J. Rubenfield et al.  
 ? TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ? TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ? FILE REFERENCE: 107196.136  
 ? CURRENT APPLICATION NUMBER: US/09/252.991A  
 ? CURRENT FILING DATE: 1999-02-18  
 ? PRIOR APPLICATION NUMBER: US 60/074,785  
 ? PRIOR FILING DATE: 1998-02-18  
 ? PRIOR APPLICATION NUMBER: US 60/094,160  
 ? PRIOR FILING DATE: 1998-07-27  
 ? NUMBER OF SEQ ID NOS: 33142  
 ? SEQ ID NO 1651  
 ? TYPE: DNA  
 ? LENGTH: 1548  
 ? ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-3651

Query Match 1.33, Score 17, DB 4, Length 1548;  
 Best Local Similarity 100.00, Pct. No. 1 26+02;  
 Matches 17, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY 422 GAAJTAGGTGGTGGAGG 438  
 ||||| ||||| ||  
 DB 320 GCAACAGGTGGTGGAGG 304

RESULT 46  
 US-08-961-527-297  
 ? Sequence 297, Application US/08961527  
 ? Patent No. 6420115  
 ? GENERAL INFORMATION:  
 ? APPLICANT: Charles Kirsch  
 ? TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
 ? NUMBER OF SEQUENCES: 391  
 ? CORRESPONDENCE ADDRESS:  
 ? ADDRESSEE: Human Genome Sciences, Inc.  
 ? STREET: 9410 K-y West Avenue  
 ? CITY: Rockville  
 ? STATE: Maryland  
 ? COUNTRY: USA  
 ? ZIP: 20850  
 ? COMPUTER READABLE FORM:  
 ? MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
 ? COMPUTER: HP Vectra 486/33  
 ? OPERATING SYSTEM: MSDOS version 6.2  
 ? SOFTWARE: ASCII Text  
 ? CURRENT APPLICATION DATA:  
 ? APPLICATION NUMBER: US/08/961.527  
 ? FILING DATE:  
 ? CLASSIFICATION: 424  
 ? PRIOR APPLICATION DATA:  
 ? APPLICATION NUMBER:  
 ? FILING DATE:  
 ? ATTORNEY/AGENT INFORMATION:  
 ? NAME: Brookes, A. Anders  
 ? REGISTRATION NUMBER: 36,373  
 ? REFERENCE/DOCKET NUMBER: PB340P1  
 ? TELECOMMUNICATION INFORMATION:  
 ? TELEPHONE: (301) 309-8504  
 ? TELEFAX: (301) 309-8512  
 ? INFORMATION FOR SEQ ID NO: 297:  
 ? SEQUENCE CHARACTERISTICS:  
 ? LENGTH: 1696 base pairs  
 ? TYPE: nucleic acid  
 ? STRANDEDNESS: double  
 ? TOPOLOGY: linear  
 US-08-961-527-297

Query Match 1.33, Score 17, DB 4, Length 1696;  
 Best Local Similarity 100.00, Pct. No. 1 26+02;  
 Matches 17, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY 1185 TATTTCATTCAGTTC 1261  
 ||||| ||||| ||  
 DB 960 TATTTCATTCAGTTC 876

RESULT 47  
 US-09-133-593A-672  
 ? Sequence 5, Application US/09133593A  
 ? Patent No. 631254  
 ? GENERAL INFORMATION:  
 ? APPLICANT: GEN. KELLER  
 ? APPLICANT: YUNG, HONG R  
 ? APPLICANT: RASHEED, USAA A  
 ? APPLICANT: RUSHEE, NY A  
 ? APPLICANT: TAD, KELLER  
 ? APPLICANT: TRUNER, ALEXANDER  
 ? TITLE OF INVENTION: TUMOR NEPHROS FACTOR RELATED RECEPTOR,  
 ? TITLE OF INVENTION: 156  
 ? FILE REFERENCE: DB 5008.2  
 ? CURRENT APPLICATION NUMBER: US/09133,593A  
 ? CURRENT FILING DATE: 1999-08-18  
 ? PRIOR APPLICATION NUMBER: 08/916,019  
 ? PRIOR FILING DATE: 1997-08-22  
 ? PRIOR APPLICATION NUMBER: 08/553,684  
 ? PRIOR FILING DATE: 1997-05-05  
 ? PRIOR APPLICATION NUMBER: 69/241,130  
 ? PRIOR FILING DATE: 1997-03-14  
 ? NUMBER OF SEQ ID NOS: 8  
 ? SOFTWARE: FASTSEQ for Windows Version 3.0  
 ? SEQ ID NO 5  
 ? TYPE: DNA  
 ? LENGTH: 1969  
 ? ORGANISM: HOMO SAPIENS  
 US-09-133-593A-672

Query Match 1.33, Score 17, DB 4, Length 1769;  
 Best Local Similarity 100.00, Pct. No. 1 26+02;  
 Matches 17, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY 1264 TTTTCTTTTCAGTTC 1220  
 ||||| ||||| ||  
 DB 1960 TTTTCTTTTCAGTTC 1046

RESULT 48  
 US-09-505-250-170  
 ? Sequence 1, Application US/09050509A  
 ? Patent No. 629148  
 ? GENERAL INFORMATION:  
 ? APPLICANT: Rosen, Glenn  
 ? APPLICANT: KAO, Peter  
 ? TITLE OF INVENTION: Synergistic Anticancer Therapy with  
 ? TITLE OF INVENTION: Cytolipolides and Death Domain Ligands  
 ? FILE REFERENCE: S00059892  
 ? CURRENT APPLICATION NUMBER: US/09/505,250A  
 ? CURRENT FILING DATE: 2000-02-15  
 ? NUMBER OF SEQ ID NOS: 4  
 ? SOFTWARE: FASTSEQ for Windows Version 3.0  
 ? SEQ ID NO 1  
 ? LENGTH: 1569  
 ? TYPE: CNA  
 ? ORGANISM: H. sapiens  
 ? FEATURES:  
 ? NAME/KEY: CDS  
 ? LOCATION: 180...1913  
 ? OTHER INFORMATION: Human TRAIL Coding Sequence  
 US-09-505-250-1

Query Match: 1.3%, Score 17; DB 4; Length 1749;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1204 TTTT-TTTT-TT-GACATA 1220  
DB 1762 TTTT-TTTT-TT-GACATA 1746

## RESULT 49

US-09-016-434-1072/C  
Sequence 1232, Application: US/09216434  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Canice Au Young  
APPLICANT: Jeffrey J. Seifhafer  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING  
TITLE OF INVENTION: PATHWAY GENE EXPRESSION  
NUMBER OF SEQUENCES: 1493  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: FAIRFAX  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/C16,414  
FILING DATE: HEREWITH

CLASSIFICATION:  
PRIOR APPLICATION DATA:

APPLICATION NUMBER:  
FILING DATE:

CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:

NAME: Zeller, Karen C.  
REGISTRATION NUMBER: 37,071

REFERENCE/DOCKET NUMBER: PA 0002 US  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 1650/855-0555  
TELEFAX: 1650/845-4166

INFORMATION FOR SEQ ID NO: 1072:  
SEQUENCE CHARACTERISTICS:

LENGTH: 1769 base pairs  
TYPE: nucleic acid

STRANDNESS: single  
TOPOLOGY: linear

IMMEDIATE SOURCE:  
LIBRARY: GENBANK

CLONE: 91149557  
US-09-016-434-1072

Query Match: 1.3%, Score 17; DB 4; Length 1769;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1204 TTTT-TTTT-TT-GACATA 1220  
DB 1762 TTTT-TTTT-TT-GACATA 1746

## RESULT 50

US-09-252-991A-16086/C  
Sequence 16086, Application: US/09212991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252.991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/374,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/394,193  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 16086  
LENGTH: 1794  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-16086

Query Match: 1.3%, Score 17; DB 4; Length 1794;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 CCGATCAGGAGAAATT 224  
DB 89 CCGATCAGGAGAAATT 73

## RESULT 51

US-09-252-991A-16516  
Sequence 16516, Application: US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252.991A

CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/374,788

PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/394,193

PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 16516  
LENGTH: 1878

TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-16516

Query Match: 1.3%, Score 17; DB 4; Length 1878;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 CCGATCAGGAGAAATT 224  
DB 1769 CCGATCAGGAGAAATT 1785

## RESULT 52

US-08-948-277A-1/C  
Sequence 1, Application: US/04949277A  
Patent No. 5849581  
GENERAL INFORMATION:

APPLICANT: Arakal, M. Catharine  
APPLICANT: Zhang, Ning

APPLICANT: Chen, Jin-Jong  
TITLE OF INVENTION: Regulators of JCP3 Gene Expression

NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:

ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
STREET: 25 DENISE DRIVE

CITY: HILLSBOROUGH  
STATE: CALIFORNIA

COUNTRY: USA  
ZIP: 94110

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/948,277A  
FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: OSMAN, RICHARD A

REGISTRATION NUMBER: 36,627

REFERENCE/DOCKET NUMBER: T97-009

TELECOMMUNICATION INFORMATION:

TELEPHONE: (650) 343-4341

TELEFAX: (650) 343-4342

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2000 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-948-277A 1

Query Match 1.38 Score 17 DB 2 Length 2000  
Best Local Similarity 100.0% Pred. No. 1.2E+02  
Matches 17 Conservative 0 Mismatches 0 Indels 0 Gaps 0

QY 1199 TTTACTTTTCTTTTGG 1215  
||||| ||||| |||||  
DB 1026 TTTACTTTTCTTTTGG 1010

RESULT 53  
US-09-169-203 1/C  
Sequence 1: Application US/09169203  
Patent No. 6576808  
GENERAL INFORMATION:  
APPLICANT: Anaral, M. Catherine  
APPLICANT: Zhang, Ning  
APPLICANT: Chen, Jin-Long  
TITLE OF INVENTION: Regulators of CYP3 Gene Expression  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
STREET: 25 DENISE DRIVE  
CITY: HILLSBOROUGH  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94010

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/169,203  
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/948,277

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: OSMAN, RICHARD A

REGISTRATION NUMBER: 36,627

REFERENCE/DOCKET NUMBER: T97-009

TELECOMMUNICATION INFORMATION:

TELEPHONE: (650) 343-4341

TELEFAX: (650) 343-4342

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2000 base pairs

TYPE: nucleic acid

STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-169-203 1

Query Match 1.38 Score 17 DB 2 Length 2000  
Best Local Similarity 100.0% Pred. No. 1.2E+02  
Matches 17 Conservative 0 Mismatches 0 Indels 0 Gaps 0

QY 1199 TTTACTTTTCTTTTGG 1215  
||||| ||||| |||||  
DB 1026 TTTACTTTTCTTTTGG 1010

RESULT 54  
US-09-319 989 5/C  
Sequence 5: Application US/09319989  
Patent No. 6583914  
GENERAL INFORMATION:  
APPLICANT: Grivell, Leslie A  
APPLICANT: Teakura De Mattos, Maureen J  
APPLICANT: Blom, Jolanda  
TITLE OF INVENTION: METHODS FOR MODULATING METABOLIC PATHWAYS OF  
TITLE OF INVENTION: MICRO-ORGANISMS AND MICRO-ORGANISMS OBTAINABLE BY SAID  
TITLE OF INVENTION: METHODS  
FILE REFERENCE: 24615-20123.00  
CURRENT APPLICATION NUMBER: US/09/319,989  
CURRENT FILING DATE: 1999-06-14  
EARLIER APPLICATION NUMBER: PCT/NL97/00049  
EARLIER FILING DATE: 1997-12-12  
EARLIER APPLICATION NUMBER: EPO 96233629  
EARLIER FILING DATE: 1996-12-12  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 5  
LENGTH: 2065  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: DNA construct  
FEATURE:  
OTHER INFORMATION: RECORDING HAP4  
NAME/KEY: CUS  
LOCATION: (201...11862)

US-09-319 989 5

Query Match 1.38 Score 17 DB 2 Length 2065  
Best Local Similarity 100.0% Pred. No. 1.2E+02  
Matches 17 Conservative 0 Mismatches 0 Indels 0 Gaps 0

QY 1199 TTTACTTTTCTTTTGG 1215  
||||| ||||| |||||  
DB 2022 TTTACTTTTCTTTTGG 2006

RESULT 55  
US-09-252-932A 4/50  
Sequence 454: Application US/09252991A  
Patent No. 6581995  
GENERAL INFORMATION:  
APPLICANT: Natic J. Rutenfeld et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO FREEDOMAS  
TITLE OF INVENTION: ABRUSINCSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.1136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 4552  
LENGTH: 2094

1 TYPE: CNA  
2 ORGANISM: Pseudomonas aeruginosa  
3 US-99 252 991A-4592

Query Match 1.33; Score 17; DB 4; Length 2094;  
Best Local Similarity 100.0%; Prod. No. 1.2e+27;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 431 GTCGAGACATGACG 447  
|||||  
DB 1119 GTCGAGACATGACG 1119

RESULT 54  
US-08-745-729-270

1 Sequence 2: Application US/08-745-729-270

2 Patent No. 5372676

3 GENERAL INFORMATION:

4 APPLICANT: Plovman, Gregory

5 APPLICANT: Mossie, Kevin

6 TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF AIR-1

7 TITLE OF INVENTION: AND/OR AIR-2 RELATED DISORDERS

8 NUMBER OF SEQUENCES: 29

9 CORRESPONDENCE ADDRESS:

10 ADDRESSEE: Lyon & Lyon

11 STREET: 613 West Fifth Street

12 CITY: Los Angeles

13 STATE: California

14 COUNTRY: U.S.A.

15 ZIP: 90071-2566

16 COMPUTER READABLE FORM:

17 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

18 MEDIUM TYPE: storage

19 COMPUTER: IBM Compatible

20 OPERATING SYSTEM: IBM P.C. DOS 5.0

21 SOFTWARE: PARTSEQ for Windows 2.0

22 CURRENT APPLICATION DATA:

23 APPLICATION NUMBER: US/08-745-729

24 FILING DATE: November 19, 1996

25 CLASSIFICATION: 530

26 PRIOR APPLICATION DATA:

27 APPLICATION NUMBER: 60/208-459

28 FILING DATE: December 19, 1996

29 APPLICATION NUMBER: 60/211,943

30 FILING DATE: August 14, 1997

31 ATTORNEY/AGENT INFORMATION:

32 NAME: Warburg, Richard J.

33 REGISTRATION NUMBER: 30,491

34 REFERENCE/DOCKET NUMBER: 210119

35 TELECOMMUNICATION INFORMATION:

36 TELEPHONE: (213) 499-1652

37 TELEFAX: (213) 955-3440

38 TELEX: 67 3510

39 INFORMATION FOR SEQ ID NO: 1:

40 SEQUENCE CHARACTERISTICS:

41 LENGTH: 2198 base pairs

42 TYPE: nucleic acid

43 STRANDEDNESS: single

44 TOPOLOGY: linear

45 MOLECULE TYPE: cDNA

46 HYPOTHEetical: NO

47 ANTI-SENSE: NO

48 US-08-745-729-2

Query Match 1.33; Score 17; DB 2; Length 2198;  
Best Local Similarity 100.0%; Prod. No. 1.2e+27;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1204 TTTTITTTTTCAGATA 1204  
|||||

DB 119: TTTTITTTTTCAGATA 2198

RESULT 57

US-08-974-655-270

1 Sequence 2: Application US/08-974-655

2 Patent No. 5372676

3 GENERAL INFORMATION:

4 APPLICANT: Plovman, Gregory

5 APPLICANT: Mossie, Kevin

6 TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF AIR-1

7 TITLE OF INVENTION: AND/OR AIR-2 RELATED DISORDERS

8 NUMBER OF SEQUENCES: 29

9 CORRESPONDENCE ADDRESS:

10 ADDRESSEE: Lyon & Lyon

11 STREET: 613 West Fifth Street

12 CITY: Los Angeles

13 STATE: California

14 COUNTRY: U.S.A.

15 ZIP: 90071-2566

16 COMPUTER READABLE FORM:

17 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

18 MEDIUM TYPE: storage

19 COMPUTER: IBM Compatible

20 OPERATING SYSTEM: IBM P.C. DOS 5.0

21 SOFTWARE: PARTSEQ for Windows 2.0

22 CURRENT APPLICATION DATA:

23 APPLICATION NUMBER: US/08-974-655

24 FILING DATE:

25 CLASSIFICATION: 435

26 PRIOR APPLICATION DATA:

27 APPLICATION NUMBER: 08/755-728

28 FILING DATE: November 25, 1996

29 APPLICATION NUMBER: 60/009,800

30 FILING DATE: December 18, 1995

31 APPLICATION NUMBER: 60/023,943

32 FILING DATE: August 14, 1996

33 ATTORNEY/AGENT INFORMATION:

34 NAME: Warburg, Richard J.

35 REGISTRATION NUMBER: 32,329

36 REFERENCE/DOCKET NUMBER: 220119

37 TELECOMMUNICATION INFORMATION:

38 TELEPHONE: (213) 499-1650

39 TELEFAX: (213) 955-3440

40 TELEX: 67-3510

41 INFORMATION FOR SEQ ID NO: 1:

42 SEQUENCE CHARACTERISTICS:

43 LENGTH: 2198 base pairs

44 TYPE: nucleic acid

45 STRANDEDNESS: single

46 TOPOLOGY: linear

47 MOLECULE TYPE: cDNA

48 HYPOTHEtical: NO

49 ANTI-SENSE: NO

50 US-08-974-655-2

Query Match 1.33; Score 17; DB 2; Length 2198;  
Best Local Similarity 100.0%; Prod. No. 1.2e+27;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1204 TTTTITTTTTCAGATA 1204  
|||||

DB 119: TTTTITTTTTCAGATA 2198

RESULT 58

US-09-261-011-270

1 Sequence 2: Application US/09-261-011

2 Patent No. 6207421

3 GENERAL INFORMATION:

4 APPLICANT: Plovman, Gregory

5 APPLICANT: Mossie, Kevin

6 TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF AIR-1

7 TITLE OF INVENTION: AND/OR AIR-2 RELATED DISORDERS

```

1 NUMBER OF SEQUENCES: 39
2 CORRESPONDENCE ADDRESS:
3 ADDRESSEE: Lyon & Lyon
4 STREET: 631 West Fifth Street
5 STREET: Suite 4700
6 CITY: Los Angeles
7 STATE: California
8 COUNTRY: U.S.A.
9 ZIP: 90071-2066
10
11 COMPUTER READABLE FORM:
12 MEDIUM TYPE: 3.5" Diskette, 1.44 MB
13 MEDIUM TYPE: storage
14 COMPUTER: IBM Compatible
15 OPERATING SYSTEM: IBM P.C. DOS 5.0
16 SOFTWARE: FASTSEQ for Windows 2.0
17 CURRENT APPLICATION DATA:
18 APPLICATION NUMBER: US/09/433,011
19 FILING DATE:
20 CLASSIFICATION:
21 PRIOR APPLICATION DATA:
22 APPLICATION NUMBER: 09/012,135
23 FILING DATE: January 22, 1998
24 APPLICATION NUMBER: 08/755,728
25 FILING DATE: No. 620401ember 25, 1996
26 APPLICATION NUMBER: 60/023,943
27 FILING DATE: August 14, 1996
28 APPLICATION NUMBER: 60/028,809
29 FILING DATE: December 28, 1995
30 ATTORNEY/AGENT INFORMATION:
31 NAME: Wareing, Richard J.
32 REGISTRATION NUMBER: 32,327
33 REFERENCE/DOCKET NUMBER: 231/282
34 TELECOMMUNICATION INFORMATION:
35 TELEPHONE: (213) 489-1630
36 TELEFAX: (213) 955-0440
37 TELEX: 673510
38 INFORMATION FOR SEQ ID NO: 2:
39 SEQUENCE CHARACTERISTICS:
40 LENGTH: 2198 base pairs
41 TYPE: nucleic acid
42 STRANDEDNESS: single
43 TOPOLOGY: linear
44 MOLECULE TYPE: cDNA
45 HYPOTHEICAL: NO
46 ANTISENSE: NO
47 US 09/283-011-2

```

```

Query Match 1.3%, Score 17; DB 3; Length 2198;
Best Local Similarity 100.0%, Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1204 TTTTCTTTTTCACATA 1220
DB 2191 TTTTCTTTTTCACATA 2175

```

```

RESULT 59
US-07-841-651-1
1 Sequence 1, Application: US/C7841651
2 Patent No.: 5410031
3 GENERAL INFORMATION:
4 APPLICANT: Fajor, Ana M.
5 APPLICANT: Wright, Ernest M.
6 TITLE OF INVENTION: Cloning and Functional Expression of a
7 TITLE OF INVENTION: Mammalian Na+/Nucleoside Cotransporter: A Member of the
8 TITLE OF INVENTION: SGLT Family
9 NUMBER OF SEQUENCES: 4
10 CORRESPONDENCE ADDRESS:
11 ADDRESSEE: Sheldon & Max
12 STREET: 225 South Lake Avenue, Ninth Floor
13 CITY: Pasadena
14 STATE: California
15 COUNTRY: USA

```

```

1 ZIP: 91101
2 COMPUTER READABLE FORM:
3 MEDIUM TYPE: 3.5" Diskette, 1.44 MB
4 COMPUTER: IBM PC compatible
5 OPERATING SYSTEM: IBM P.C. DOS/MS-DOS
6 SOFTWARE: Patentlic Release #1.0, Version #1.25
7 CURRENT APPLICATION DATA:
8 APPLICATION NUMBER: US/07/841,651
9 FILING DATE: 19920224
10 CLASSIFICATION: 435
11 ATTORNEY/AGENT INFORMATION:
12 NAME: Vaseco, Sarajynn
13 REGISTRATION NUMBER: 11,953
14 REFERENCE/DOCKET NUMBER: 9772
15 TELECOMMUNICATION INFORMATION:
16 TELEPHONE: (818) 340-4440
17 TELEFAX: (818) 340-4440
18 INFORMATION FOR SEQ ID NO: 1:
19 SEQUENCE CHARACTERISTICS:
20 LENGTH: 2238 base pairs
21 TYPE: nucleic acid
22 STRANDEDNESS: double
23 TOPOLOGY: linear
24 MOLECULE TYPE: cDNA
25 HYPOTHEICAL: NO
26 ANTISENSE: NO
27 ORIGINAL SOURCE:
28 ORGANISM: Cystelasma emiculus
29 FEATURE:
30 NAME/KEY: cDS
31 LOCATION: 7..2002
32 US 07-841-651-1

```

```

Query Match 1.3%, Score 17; DB 3; Length 2238;
Best Local Similarity 100.0%, Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 431 GCTGAGAGGTGCTGGAGG 447
DB 1923 GCTTAAAGAAAGAGAG 1939

```

```

RESULT 60
US-09-252-951A-3510
1 Sequence 3620, Application: US/09252951A
2 Patent No.: 6551796
3 GENERAL INFORMATION:
4 APPLICANT: Mair, J. Robertfield et al.
5 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
6 TITLE OF INVENTION: ARABINOSA FOR DIAGNOSTICS AND THERAPEUTICS
7 FILE REFERENCE: 107196,116
8 CURRENT APPLICATION NUMBER: US/09/252,951A
9 CURRENT FILING DATE: 1999-02-18
10 PRIOR APPLICATION NUMBER: US 60/094,784
11 PRIOR FILING DATE: 1998-07-15
12 PRIOR APPLICATION NUMBER: US 60/094,194
13 PRIOR FILING DATE: 1998-07-27
14 NUMBER OF SEQ ID NOS: 3,142
15 SEQ ID NO 9510
16 LENGTH: 2253
17 TYPE: cDNA
18 ORGANISM: Pseudomonas aeruginosa
19 US-09-252-951A 3510

```

```

Query Match 1.3%, Score 17; DB 4; Length 2253;
Best Local Similarity 100.0%, Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 422 GCAGCAGGTGCTGGAGG 438
DB 1959 GCAGCAGGTGCTGGAGG 1975

```



RESULT 61  
 US-08-156-253 1  
 ? Sequence 1: Application US/09156253C  
 ? Patent No. 6001652  
 ? GENERAL INFORMATION:  
 ? APPLICANT: Monia, Brett P.  
 ? APPLICANT: Cowart, Brenda F.  
 ? APPLICANT: Baker, Dex M.  
 ? TITLE OF INVENTION: Antisense Modulation of cREL Expression  
 ? FILE REFERENCE: RTS-0010  
 ? CURRENT APPLICATION NUMBER: US/09/156-253  
 ? CURRENT FILING DATE: 1998-09-15  
 ? NUMBER OF SEQ ID NOS: 48  
 ? SOFTWARE: Patent In Ver. 2.0  
 ? SEQ ID NO 1:  
 ? LENGTH: 2337  
 ? TYPE: DNA  
 ? ORGANISM: Homo sapiens  
 ? FEATURES:  
 ? NAME/KEY: unsure  
 ? LOCATION: 411  
 ? NAME/KEY: unsure  
 ? LOCATION: 2277  
 ? US-08-156-253 1

Query Match 1.38; Score 17; DB 3; Length 2337;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 780 CTGTGATATGATGACA 905  
 |||||  
 DB 166 CTGTGATATGATGACA 169

RESULT 62  
 US-08-156-552-1  
 ? Sequence 1: Application US/0806852  
 ? Patent No. 6000016  
 ? GENERAL INFORMATION:  
 ? APPLICANT: Edwards, Robert H.  
 ? TITLE OF INVENTION: Vesicle Membrane Transport Proteins  
 ? NUMBER OF SEQUENCES: 17  
 ? CORRESPONDENCE ADDRESS:  
 ? ADDRESSEE: Sheldon & Mak  
 ? STREET: 225 South Lake Avenue, Ninth Floor  
 ? CITY: Pasadena  
 ? STATE: California  
 ? COUNTRY: USA  
 ? ZIP: 91101  
 ? COMPUTER READABLE FORM:  
 ? MEDIUM TYPE: floppy disk  
 ? COMPUTER: IBM PC compatible  
 ? OPERATING SYSTEM: PC-DOS/MS-DOS  
 ? SOFTWARE: Patent In Release #1.0, Version #1.25  
 ? CURRENT APPLICATION DATA:  
 ? APPLICATION NUMBER: US/08/063-552  
 ? FILING DATE: 19930514  
 ? CLASSIFICATION: S30  
 ? ATTORNEY/AGENT INFORMATION:  
 ? NAME: Farber, Michael B.  
 ? REGISTRATION NUMBER: 32,612  
 ? REFERENCE/DOCKET NUMBER: 9067 1  
 ? TELEPHONE: (818) 796-4030  
 ? TELEFAX: (818) 795-6321  
 ? INFORMATION FOR SEQ ID NO: 1:  
 ? SEQUENCE CHARACTERISTICS:  
 ? LENGTH: 2421 base pairs  
 ? TYPE: NUCLEIC ACID  
 ? STRANDEDNESS: double  
 ? TOPOLOGY: linear  
 ? MOLECULE TYPE: cDNA

? HYPOTHETICAL: NO  
 ? ANTI-SENSE: NO  
 ? ORIGINAL SOURCE:  
 ? ORGANISM: Rattus rattus  
 ? FEATURES:  
 ? NAME/KEY: CDS  
 ? LOCATION: 262...1924  
 ? US-08-063-552-1

Query Match 1.38; Score 17; DB 1; Length 2421;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1204 TTTTCTTTTGGACATA 1223  
 |||||  
 DB 2193 TTTTCTTTTGGACATA 2199

RESULT 63  
 PCT-US93-05704-1  
 ? Sequence 1: Application PCT/US9305704  
 ? GENERAL INFORMATION:  
 ? APPLICANT: Edwards, Robert H.  
 ? TITLE OF INVENTION: Vesicle Membrane Transport Proteins  
 ? NUMBER OF SEQUENCES: 17  
 ? CORRESPONDENCE ADDRESS:  
 ? ADDRESSEE: Sheldon & Mak  
 ? STREET: 225 South Lake Avenue, Ninth Floor  
 ? CITY: Pasadena  
 ? STATE: California  
 ? COUNTRY: USA  
 ? ZIP: 91101  
 ? COMPUTER READABLE FORM:  
 ? MEDIUM TYPE: floppy disk  
 ? COMPUTER: IBM PC compatible  
 ? OPERATING SYSTEM: PC-DOS/MS-DOS  
 ? SOFTWARE: Patent In Release #1.0, Version #1.25  
 ? CURRENT APPLICATION DATA:  
 ? APPLICATION NUMBER: PCT/US93/05704  
 ? FILING DATE: 19930611  
 ? CLASSIFICATION:  
 ? ATTORNEY/AGENT INFORMATION:  
 ? NAME: Farber, Michael B.  
 ? REGISTRATION NUMBER: 32,612  
 ? REFERENCE/DOCKET NUMBER: 9067-1PCT  
 ? TELEPHONE: (818) 796-4030  
 ? TELEFAX: (818) 795-6321  
 ? INFORMATION FOR SEQ ID NO: 1:  
 ? SEQUENCE CHARACTERISTICS:  
 ? LENGTH: 2421 base pairs  
 ? TYPE: NUCLEIC ACID  
 ? STRANDEDNESS: double  
 ? TOPOLOGY: linear  
 ? MOLECULE TYPE: cDNA  
 ? HYPOTHETICAL: NO  
 ? ANTI-SENSE: NO  
 ? ORIGINAL SOURCE:  
 ? ORGANISM: Rattus rattus  
 ? FEATURES:  
 ? NAME/KEY: CDS  
 ? LOCATION: 262...1924  
 ? PCT-US93-05704-1

Query Match 1.38; Score 17; DB 5; Length 2421;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1204 TTTTCTTTTGGACATA 1220  
 |||||  
 DB 2193 TTTTCTTTTGGACATA 2199

## RESULT 64

US-09-620-312D-223  
 ? Sequence 223, Application US/09620312D  
 ? Patent No. 6569662  
 ? GENERAL INFORMATION:  
 ? APPLICANT: Tang, Y. Tom  
 ? APPLICANT: Liu, Chenghua  
 ? APPLICANT: Asundi, Vinod  
 ? APPLICANT: Zhang, Jie  
 ? APPLICANT: Ren, Feiyang  
 ? APPLICANT: Chen, Rui-hong  
 ? APPLICANT: Zhao, Qing A.  
 ? APPLICANT: Wehrman, Tom  
 ? APPLICANT: Xue, Aidong J.  
 ? APPLICANT: Yang, Yonghong  
 ? APPLICANT: Wang, Jian-Rui  
 ? APPLICANT: Zhou, Ping  
 ? APPLICANT: Ma, Yurqing  
 ? APPLICANT: Wang, Dunrui  
 ? APPLICANT: Wang, Zhiwei  
 ? APPLICANT: John Tillinghast  
 ? APPLICANT: Drmanac, Radoje T.  
 ? TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
 ? FILE REFERENCE: 284C182B  
 ? CURRENT APPLICATION NUMBER: US/09/620,312D  
 ? PRIOR FILING DATE: 2000-07-19  
 ? PRIOR APPLICATION NUMBER: 09/552,317  
 ? PRIOR FILING DATE: 2000-04-25  
 ? PRIOR APPLICATION NUMBER: 09/489,725  
 ? PRIOR FILING DATE: 2000-01-21  
 ? NUMBER OF SEQ ID NOS: 1103  
 ? SOFTWARE: pc\_Fl\_genes Version 1.0  
 ? SEQ ID NO 223  
 ? LENGTH: 2457  
 ? TYPE: DNA  
 ? ORGANISM: Homo sapiens  
 ? FEATURES:  
 ? NAME/KEY: COS  
 ? LOCATION: (175)..(1949)  
 US-09-620-312D-223

Query Match 1.31, Score 17, DB 4, Length 2457  
 Best Local Similarity 100.0%, Pred. No. 1103/223  
 Matches 17, Conservation 0, Mismatches 0, Indels 0, Gaps 0:

QY 789 CTGTGAATATGATGACA 805  
 |||.....|  
 DB 1562 CTGTGAATATGATGACA 1578

## RESULT 65

US-09-620-312D-547/C  
 ? Sequence 547, Application US/09620312D  
 ? Patent No. 6569662  
 ? GENERAL INFORMATION:  
 ? APPLICANT: Tang, Y. Tom  
 ? APPLICANT: Liu, Chenghua  
 ? APPLICANT: Asundi, Vinod  
 ? APPLICANT: Zhang, Jie  
 ? APPLICANT: Ren, Feiyang  
 ? APPLICANT: Chen, Rui-hong  
 ? APPLICANT: Zhao, Qing A.  
 ? APPLICANT: Wehrman, Tom  
 ? APPLICANT: Xue, Aidong J.  
 ? APPLICANT: Yang, Yonghong  
 ? APPLICANT: Wang, Jian-Rui  
 ? APPLICANT: Zhou, Ping  
 ? APPLICANT: Ma, Yurqing  
 ? APPLICANT: Wang, Dunrui  
 ? APPLICANT: Wang, Zhiwei  
 ? APPLICANT: John Tillinghast  
 ? APPLICANT: Drmanac, Radoje T.

? TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
 ? FILE REFERENCE: 284C182B  
 ? CURRENT APPLICATION NUMBER: US/09/620,312D  
 ? PRIOR FILING DATE: 2000-07-19  
 ? PRIOR APPLICATION NUMBER: 09/552,317  
 ? PRIOR FILING DATE: 2000-04-25  
 ? PRIOR APPLICATION NUMBER: 09/489,725  
 ? PRIOR FILING DATE: 2000-01-21  
 ? NUMBER OF SEQ ID NOS: 1103  
 ? SOFTWARE: pc\_Fl\_genes Version 1.0  
 ? SEQ ID NO 547  
 ? LENGTH: 2660  
 ? TYPE: DNA  
 ? ORGANISM: Homo sapiens  
 ? FEATURES:  
 ? NAME/KEY: COS  
 ? LOCATION: (175)..(1949)  
 US-09-620-312D-547

Query Match 1.31, Score 17, DB 4, Length 2660  
 Best Local Similarity 100.0%, Pred. No. 1103/223  
 Matches 17, Conservation 0, Mismatches 0, Indels 0, Gaps 0:

QY 1193 TTTATTTTCTTTCTTCTG 1215  
 |||.....|  
 DB 1697 TTTATTTTCTTTCTTCTG 1681

## RESULT 66

US-09-854-856-51/C  
 ? Sequence 53, Application US/09854856  
 ? Patent No. 6541252  
 ? GENERAL INFORMATION:  
 ? APPLICANT: Walke, D. Wade  
 ? APPLICANT: Hilbun, Elin  
 ? APPLICANT: Donoto, Gregory  
 ? APPLICANT: Turner, C. Alexander Jr.  
 ? TITLE OF INVENTION: No. 6541252el Run in Kinases and Polynucleotides  
 ? FILE REFERENCE: Enclosing the Same  
 ? CURRENT APPLICATION NUMBER: US/09/854,856  
 ? PRIOR FILING DATE: 2000-05-14  
 ? PRIOR APPLICATION NUMBER: US/09/206,618  
 ? PRIOR FILING DATE: 2000-03-19  
 ? NUMBER OF SEQ ID NOS: 24  
 ? SOFTWARE: FastSeq for Windows Version 4.0  
 ? SEQ ID NO 53  
 ? LENGTH: 2685  
 ? TYPE: DNA  
 ? ORGANISM: Homo sapiens  
 US-09-854-856-53

Query Match 1.31, Score 17, DB 4, Length 2685  
 Best Local Similarity 100.0%, Pred. No. 1103/223  
 Matches 17, Conservation 0, Mismatches 0, Indels 0, Gaps 0:

QY 529 TGTGAGGAAATGATGACA 545  
 |||.....|  
 DB 2370 TGTGAGGAAATGATGACA 2384

## RESULT 67

US-09-854-856-57/C  
 ? Sequence 37, Application US/09854856  
 ? Patent No. 6541252  
 ? GENERAL INFORMATION:  
 ? APPLICANT: Walke, D. Wade  
 ? APPLICANT: Hilbun, Elin  
 ? APPLICANT: Donoto, Gregory  
 ? APPLICANT: Turner, C. Alexander Jr.  
 ? TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides  
 ? FILE REFERENCE: Enclosing the Same

FILE REFERENCE: LEX-0178-USA  
 CURRENT APPLICATION NUMBER: US/09/854,856  
 CURRENT FILING DATE: 2001-05-14  
 PRIOR APPLICATION NUMBER: US 60/206,015  
 PRIOR FILING DATE: 2000-05-19  
 NUMBER OF SEQ ID NOS: 64  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 17  
 LENGTH: 2769  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 US-09-854-856-5

Query Match  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02; Length 2769;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 529 TGAAGCCAACTGAGCA 545  
 |||||  
 DB 2454 TGAAGCCAACTGAGCA 2438

RESULT 69  
 US-09-854-856-21/c  
 Sequence 21, Application US/09854856  
 Patent No. 6541252  
 GENERAL INFORMATION:  
 APPLICANT: Walke, D. Wade  
 APPLICANT: Hilborn, Erin  
 APPLICANT: Donoho, Gregory  
 APPLICANT: Turner, C. Alexander Jr.  
 TITLE OF INVENTION: No. 6541252: Human Kinases and Polynucleotides  
 TITLE OF INVENTION: Encoding the Same  
 FILE REFERENCE: LEX-0178-USA  
 CURRENT APPLICATION NUMBER: US/09/854,856  
 CURRENT FILING DATE: 2001-05-14  
 PRIOR APPLICATION NUMBER: US 60/206,015  
 PRIOR FILING DATE: 2000-05-19  
 NUMBER OF SEQ ID NOS: 64  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 21  
 LENGTH: 2865  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 US-09-854-856-21

Query Match  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02; Length 2865;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 529 TGAAGCCAACTGAGCA 545  
 |||||  
 DB 2454 TGAAGCCAACTGAGCA 2438

RESULT 69  
 US-09-854-856-5/c  
 Sequence 5, Application US/09854856  
 Patent No. 6541252  
 GENERAL INFORMATION:  
 APPLICANT: Walke, D. Wade  
 APPLICANT: Hilborn, Erin  
 APPLICANT: Donoho, Gregory  
 APPLICANT: Turner, C. Alexander Jr.  
 TITLE OF INVENTION: No. 6541252: Human Kinases and Polynucleotides  
 TITLE OF INVENTION: Encoding the Same  
 FILE REFERENCE: LEX-0178-USA  
 CURRENT APPLICATION NUMBER: US/09/854,856  
 CURRENT FILING DATE: 2001-05-14  
 PRIOR APPLICATION NUMBER: US 60/206,015  
 PRIOR FILING DATE: 2000-05-19  
 NUMBER OF SEQ ID NOS: 64  
 SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 5  
 LENGTH: 2949  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 US-09-854-856-5

Query Match  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02; Length 2949;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 529 TGAAGCCAACTGAGCA 545  
 |||||  
 DB 2634 TGAAGCCAACTGAGCA 2618

RESULT 70  
 US-09-319-989-9/c  
 Sequence 9, Application US/09319999  
 Patent No. 6190914  
 GENERAL INFORMATION:  
 APPLICANT: Grivell, Leslie A.  
 APPLICANT: Bloem, Coisanda  
 APPLICANT: Teixeira De Matos, Maarten J.  
 TITLE OF INVENTION: METHODS FOR MODULATING METABOLIC PATHWAYS OF  
 TITLE OF INVENTION: MICRO-ORGANISMS AND MICRO-ORGANISMS OBTAINABLE BY SAID  
 TITLE OF INVENTION: METHODS  
 FILE REFERENCE: 24615-20123-00  
 CURRENT APPLICATION NUMBER: US/09/319,989  
 CURRENT FILING DATE: 1999-06-14  
 EARLIER APPLICATION NUMBER: PCT/NL97/00688  
 EARLIER FILING DATE: 1997-12-12  
 EARLIER APPLICATION NUMBER: EPC 96203520  
 EARLIER FILING DATE: 1996-12-12  
 NUMBER OF SEQ ID NOS: 9  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 9  
 LENGTH: 3081  
 TYPE: DNA  
 ORGANISM: Artificial Sequence  
 FEATURE:  
 OTHER INFORMATION: Description of Artificial Sequence: DNA construct consisting  
 OTHER INFORMATION: of the TEF2 promoter fused to the coding region of HAP4  
 US-09-319-989-9

Query Match  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02; Length 3081;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1199 TTACTTTTCTTTTCTG 1215  
 |||||  
 DB 2261 TTACTTTTCTTTTCTG 2239

RESULT 71  
 US-09-620-3120-713  
 Sequence 713, Application US/096203120  
 Patent No. 6569662  
 GENERAL INFORMATION:  
 APPLICANT: Tang, Y. Tom  
 APPLICANT: Liu, Chonghua  
 APPLICANT: Asundi, Vinod  
 APPLICANT: Zhang, Jie  
 APPLICANT: Ren, Feiyen  
 APPLICANT: Chen, Rui-Hong  
 APPLICANT: Zhao, Qing A.  
 APPLICANT: Wehrman, Tom  
 APPLICANT: Xue, Aidong J.  
 APPLICANT: Yang, Yonghong  
 APPLICANT: Wang, Chao-Rui  
 APPLICANT: Zhou, Ping  
 APPLICANT: Ma, Yunqing  
 APPLICANT: Wang, Dunrui  
 APPLICANT: Wang, Zhiwei

? APPLICANT: John Tillinghast  
 ? APPLICANT: Drmanac, Radomir T.  
 ? TITLE OF INVENTION: No. 6596622: Nucleic Acids and  
 ? FILE REFERENCE: Polypeptides  
 ? CURRENT APPLICATION NUMBER: US/09/620,312D  
 ? PRIOR FILING DATE: 2000-07-19  
 ? PRIOR APPLICATION NUMBER: 09/552,317  
 ? PRIOR FILING DATE: 2000-04-25  
 ? PRIOR APPLICATION NUMBER: 09/486,725  
 ? PRIOR FILING DATE: 2000-01-21  
 ? NUMBER OF SEQ ID NOS: 1105  
 ? SOFTWARE: PL FL genes Version 1.0  
 ? SEQ ID NO 713  
 ? LENGTH: 3169  
 ? TYPE: DNA  
 ? ORGANISM: Homo sapiens  
 ? FEATURE:  
 ? NAME/KEY: CDS  
 ? LOCATION: 1161..12146;  
 ? US-09-620-312D-713

Query Match 1.3%; Score 17; DB 4; Length 3169;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1199 TTTACTTTTITTTTITG 1215  
 DB 2559 TTTACTTTTITTTITG 2575

RESULT 72  
 ? Sequence 1, Application US/09286959B  
 ? Patent No. 6300131  
 ? GENERAL INFORMATION:  
 ? APPLICANT: Johns Hopkins University  
 ? APPLICANT: Greider, Carol W.  
 ? APPLICANT: Li, Siyuan  
 ? TITLE OF INVENTION: TELOMERASE-ASSOCIATED PROTEINS  
 ? FILE REFERENCE: 07265/157001  
 ? CURRENT APPLICATION NUMBER: US/09/286,959B  
 ? CURRENT FILING DATE: 1999-04-06  
 ? PRIOR APPLICATION NUMBER: 60/080,783  
 ? PRIOR FILING DATE: 1998-04-06  
 ? NUMBER OF SEQ ID NOS: 24  
 ? SOFTWARE: FastSeq for Windows Version 4.0  
 ? SEQ ID NO 1  
 ? LENGTH: 3190  
 ? TYPE: DNA  
 ? ORGANISM: Homo sapiens  
 ? US-09-286-959B-1

Query Match 1.3%; Score 17; DB 4; Length 3190;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1105 AACTGTCCTGGGTACT 1125  
 DB 2613 AACTGTCCTGGGTACT 2597

RESULT 73  
 ? Sequence 21, Application US/08565762  
 ? Patent No. 6280463  
 ? GENERAL INFORMATION:  
 ? APPLICANT: Kottir, Vignar  
 ? APPLICANT: Gavrilas, Victoria  
 ? TITLE OF INVENTION: ESSENTIAL FUNGAL GENES AND THEIR USE  
 ? FILE REFERENCE: 07334/062001  
 ? CURRENT APPLICATION NUMBER: US/08/965,762  
 ? CURRENT FILING DATE: 1997-11-07

? NUMBER OF SEQ ID NOS: 10  
 ? SOFTWARE: FastSeq for Windows Version 1.0  
 ? SEQ ID NO 21  
 ? LENGTH: 3343  
 ? TYPE: DNA  
 ? ORGANISM: Saccharomyces cerevisiae  
 ? US-08-965-762-21

Query Match 1.3%; Score 17; DB 3; Length 3343;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 721 TTTATACATCTTTTGA 787  
 DB 2637 TTTATACATCTTTTGA 2621

RESULT 74  
 ? Sequence 8, Application US/0942998A  
 ? Patent No. 6190914  
 ? GENERAL INFORMATION:  
 ? APPLICANT: Grivell, Leslie A.  
 ? APPLICANT: Telkella De Vries, Maarten J.  
 ? APPLICANT: Scom, Oslandi

? TITLE OF INVENTION: METHODS FOR MODULATING METABOLIC PATHWAYS OF  
 ? TITLE OF INVENTION: MICRO ORGANISMS AND MICRO ORGANISMS OBTAINABLE BY SAID  
 ? TITLE OF INVENTION: METHODS  
 ? FILE REFERENCE: 24615-20121-00  
 ? CURRENT APPLICATION NUMBER: US/09/419,993  
 ? CURRENT FILING DATE: 1999-06-14  
 ? EARLIER APPLICATION NUMBER: PCT/NL47/00648  
 ? EARLIER FILING DATE: 1997-12-12  
 ? EARLIER APPLICATION NUMBER: EP 062003.0  
 ? EARLIER FILING DATE: 1997-12-12  
 ? NUMBER OF SEQ ID NOS: 1  
 ? SOFTWARE: Patent In Ver. 2.0  
 ? SEQ ID NO 8  
 ? LENGTH: 3360  
 ? TYPE: DNA  
 ? ORGANISM: Artificial Sequence  
 ? FEATURE:  
 ? OTHER INFORMATION: Description of Artificial Sequence: DNA construct consisting  
 ? OTHER INFORMATION: of the GPOL protein fused to the coding region of HA94  
 ? US-09-319-993-8

Query Match 1.3%; Score 17; DB 3; Length 3360;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1199 TTTACTTTTITTTTITG 1215  
 DB 2630 TTTACTTTTITTTTITG 2614

RESULT 75  
 ? Sequence 2, Application US/08457642A  
 ? Patent No. 5837524  
 ? GENERAL INFORMATION:  
 ? APPLICANT: Sina Lov  
 ? APPLICANT: Josep Sullensinger  
 ? TITLE OF INVENTION: PWR RELATED BR 0015  
 ? TITLE OF INVENTION: AND METHODS  
 ? NUMBER OF SEQUENCES: 16  
 ? CORRESPONDENCE ADDRESS:  
 ? ADDRESSEE: Lyon & Lyon  
 ? STREET: 633 West Fifth Street  
 ? STREET: Suite 4700  
 ? CITY: Los Angeles  
 ? STATE: California  
 ? COUNTRY: U.S.A.  
 ? ZIP: 90071-2066

```

1  COMPUTER READABLE FORM:
2  MEDIUM TYPE: 3.5" Diskette, 1.44 Kb
3  MEDIUM TYPE: Storage
4  COMPUTER: IBM Compatible
5  OPERATING SYSTEM: IBM PC DOS 5.0
6  SOFTWARE: Word Perfect 5.1
7  CURRENT APPLICATION DATA:
8  APPLICATION NUMBER: US/06/437,42A
9  FILING DATE: December 15, 1994
10 CLASSIFICATION: 516
11 PRIOR APPLICATION DATA:
12 APPLICATION NUMBER:
13 FILING DATE:
14 ATTORNEY/AGENT INFORMATION:
15 NAME: Warburg, Richard J
16 REGISTRATION NUMBER: 21,123
17 REFERENCE/DOCKET NUMBER: 201/010
18 TELECOMMUNICATION INFORMATION:
19 TELEPHONE: (213) 489-1600
20 TELEFAX: (213) 955-0440
21 TELEX: 67-3510
22 INFORMATION FOR SEQ ID NO: 2:
23 SEQUENCE CHARACTERISTICS:
24     LENGTH: 1416
25     TYPE: nucleic acid
26     STRANDEDNESS: single
27     TOPOLOGY: linear
28     MOLECULE TYPE: nucleic
29 US 58-357 642A 2
30
31 Query March 1994, Accession 17, DB 2, Length 1416
32 Best Local Similarity 100.0%, Prev. No. 1,254,027
33 Matches 17, Conservative 0, Mismatches 0, Indels 0, Gaps 0:
34
35 CY 117 AACGCTTCAGCAGTAC 137
36      ||| ||||| |||
37 CB 803 AACGCTTCAGCAGTAC 811

```

Search completed: October 21, 2003, 06:45:00  
 Job time: 156 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

OX nucleic acid search, using sw model

Run on: October 21, 2003, 05:16:37 / Search time 5026 Seconds  
(without alignment)  
10426.927 Million cell updates/sec

Title: US-11-029-137-1

Perfect score: 1281

Sequence: 1 codatagcgcgacccatataaaacattggatgaatgaactg 1281

Scoring table: Gapped NUC

Scores: Supx=50.0 , Gapext=60.0

Searched: 2888711 seqs, 2045481356 residues

Word size : 9

Total number of hits satisfying chosen parameters: 5777422

Minimum FR seq length: 5

Maximum FR seq length: 2000000000

Post processing: Listing first 100 summaries

Database : GenBank :

	No.	Score	Query Match	Length	ID	Description
C	1	1281	100.0	1363	9	AF158839 Homo sapi
C	2	1281	100.0	1303	9	AF121617 Homo sapi
C	3	1011	78.9	1011	6	AX017517 Sequence
C	4	1011	78.9	1011	6	AX017517 Human nuc
C	5	1005	78.9	1011	6	AX015050 Sequence
C	6	1004	78.4	1033	9	BC011936 Homo sapi
C	7	888	69.3	1733.0	9	AC006160 Homo sapi
C	8	857	66.9	882	6	AX014817 Sequence
C	9	776	60.7	822	6	B015341 Human tto
C	10	667	52.1	1063	9	AF117679 Homo sapi
C	11	456	38.7	789	9	AF117680 Homo sapi
C	12	404	31.5	901	9	AF117678 Homo sapi
C	13	305	23.6	467	6	BC108605 EST and e
C	14	253	19.8	465	4	BC025539 Sequence
C	15	56	4.4	904	10	BC024941 Mus muscu
C	16	56	4.4	21762	2	AC092771 Mus muscu
C	17	40	3.1	127188	2	AC005063 Homo sapi
C	18	40	3.1	121384	9	AL135984 Human RNA
C	19	39	3.0	55304	9	HS227LS
C	20	37	2.9	173911	9	AL334992 Human DNA
C	21	34	2.7	119561	9	AL138799 Human DNA
C	22	32	2.5	127867	9	HS694B1A Human DNA
C	23	29	2.3	108444	9	AP000725 Homo sapi
C	24	28	2.2	163456	9	AC017071 Homo sapi
C	25	27	2.1	233422	2	AC098358 Rattus no
C	26	27	2.1	246937	2	AC121198 Rattus no
C	27	23	1.9	106293	9	AP006259 Homo sapi
C	28	23	1.8	167446	2	AC137718 Homo sapi
C	29	23	1.8	134686	9	AC079296 Homo sapi
C	30	23	1.8	132203	2	AC026077 Homo sapi
C	31	23	1.8	243249	2	AC097383 Rattus no
C	32	23	1.8	280190	2	AC111223 Rattus no
C	33	22	1.7	57608	2	AC105248 Homo sapi
C	34	22	1.7	61859	2	AC090374 Homo sapi
C	35	22	1.7	61859	2	AC090374 Homo sapi
C	36	22	1.7	64681	2	AC105143 Homo sapi
C	37	22	1.7	64693	2	AC105143 Homo sapi
C	38	22	1.7	110200	2	AC079625 Homo sapi
C	39	22	1.7	110200	2	AC079625 Homo sapi
C	40	22	1.7	126284	9	Continuation of
C	41	22	1.7	143564	9	AC012254 Homo sapi
C	42	22	1.7	143564	9	AC012254 Homo sapi
C	43	22	1.7	151073	2	AC144823 Mus muscu
C	44	22	1.7	160861	2	AC144823 Mus muscu
C	45	22	1.7	161783	2	BX005421 Dario rer
C	46	22	1.7	162382	2	AC018861 Homo sapi
C	47	22	1.7	185576	2	AC110390 Homo sapi
C	48	22	1.7	202103	9	AC101016 Homo sapi
C	49	22	1.7	203469	2	AC115853 Mus muscu
C	50	22	1.7	203469	2	AC115853 Mus muscu
C	51	22	1.7	206810	2	BX470141 Dario rer
C	52	22	1.7	227588	2	AC087162 Mus -JUSC
C	53	22	1.7	227695	2	AC115252 Rattus no
C	54	22	1.7	229678	2	AC103232 Rattus no
C	55	22	1.7	244620	2	AC011176 Mus muscu
C	56	22	1.7	347310	2	AC018743 Homo sapi
C	57	21	1.6	7735	11	BV061249 S21pf6193
C	58	21	1.6	10407	1	AEO06053 Pasteurel
C	59	21	1.6	25290	2	AC017669 Drosophil
C	60	21	1.6	30190	9	AC108010 Homo sapi
C	61	21	1.6	55406	2	AC115980 Mus muscu
C	62	21	1.6	76848	9	AC003041 Homo sapi
C	63	21	1.6	91469	9	AC010444 Homo sapi
C	64	21	1.6	107685	9	AC122698 Homo sapi
C	65	21	1.6	111682	9	AC115612 Dictyoste
C	66	21	1.6	114945	9	AL021818 Human DNA

Died. NO. is the number of results predicted by BLAST to have a

Printed No. is the number of results predicted by chance to have a

```
66 1.6 118759 9 AC129917 Homo sapi:
67 1.6 120392 9 AP005548 Oryza sat
68 1.6 121922 9 HS3927M24 Human DNA
69 1.6 127573 2 AP005755 Oryza sat
70 1.6 129497 2 AC121506 Mus muscu
71 1.6 137289 9 AC022828 Homo sapi
72 1.6 139030 2 AC159533 Mus muscu
73 1.6 139517 9 AC159533 Mus muscu
74 1.6 143244 9 AC091997 Homo sapi
75 1.6 145153 9 AL590031 Human DNA
76 1.6 150069 2 AC041042 Homo sapi
77 1.6 156761 2 AP053312 Oryza sat
78 1.6 162215 9 AC004896 Homo sapi
79 1.6 164076 9 AC021059 Homo sapi
80 1.6 165159 2 AC068577 Homo sapi
81 1.6 168779 9 AC026405 Homo sapi
82 1.6 170293 9 AC090940 Homo sapi
83 1.6 176976 2 AL331095 Homo sapi
84 1.6 179270 9 AC134580 Homo sapi
85 1.6 180704 2 AC116513 Homo sapi
86 1.6 181343 10 AL671882 Mouse DNA
87 1.6 184964 2 AP055702 Oryza sat
88 1.6 185839 2 AC023278 Homo sapi
89 1.6 192070 9 AC090616 Homo sapi
90 1.6 194121 10 AL331842 Mouse DNA
91 1.6 194308 3 AC009377 Drosophila
92 1.6 199025 2 AC073515 Homo sapi
93 1.6 205720 2 AC107864 Mus muscu
94 1.6 206937 5 BX001018 Zebrafish
95 1.6 207411 10 AC114916 Mus muscu
96 1.6 215357 2 AC132379 Mus muscu
97 1.6 226460 10 AC098720 Mus muscu
98 1.6 229669 10 AC116996 Mus muscu
99 1.6 231291 2 AC115750 Mus muscu
```

## ALIGNMENTS

```
RESULT 1
AF358829 1363 bp mRNA linear PRI 11-JAN-2002
LOCUS AF358829 Homo sapiens tumor angiogenesis marker mRNA, complete cds.
DEFINITION AF358829
VERSION AF358829.1 GI:13605459
KEYWORDS Homo sapiens (human)
```

```
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1363)
AUTHORS Liu, C., Zhang, L., Shao, Z. M., Beatty, P., Saitippour, M., Lane, T. F.,
Basky, S. H., Livingston, E. and Nguyen, M.
TITLE Identification of a novel endothelial-derived gene EG-1
JOURNAL Biochem. Biophys. Res. Commun. 290 (1), 602-612 (2002)
MEDLINE 21538037
PUBMED 11779215
REFERENCE 2 (bases 1 to 1363)
AUTHORS Nguyen, M. H. and Liu, C.
TITLE A Novel Gene as Tumor Angiogenesis Marker
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1363)
AUTHORS Nguyen, M. H. and Liu, C.
TITLE Direct Submission
JOURNAL Submitted (08-MAR-2003) Surgical Oncology, UCLA Medical School,
16833 Le Conte Ave., Los Angeles, CA 90095, USA
FEATURES
source
1..1363
/organism="Homo sapiens"
/db_xref="taxon:9606"
13..549
```

```
/random start=1
/product="tumor angiogenesis marker"
/procid.id="AAK3274.1"
/db_xref="GI:13605459"
/translation="MAAPHSQVFSSQPPQAPQPLGPGQASLLQAAKAPKAPFSSSTL
VQHLSESPACAFASVUSCVYNTVQEEIFETGDCIOKFLDIARCTECPFLQKELQ
SVQKEQVKEVSELSERNEKHKQKALVUKLTKLHWQVLEIDINQKAPACIFQUS
LAYLEVAFAKTPAPKPT"
PASE COUNT 410 a 245 c 305 g
ORIGIN
Query Match: 100.0% Score 108.1 DB 9; Length 1363;
Best Local Similarity: 100.0% Pident No. 9;
Matches 128; Conservation: 0; Mismatches 0; Indels 0; Gaps 0;
CY 1 GAGACATATGCGTTCATAGAGGAGGATTTTTCAGGACACACCGGTCCTTCAGAC 60
DB 1 GAGACATATGCGTTCATAGAGGAGGATTTTTCAGGACACACCGGTCCTTCAGAC 60
CY 2 GTTCACACAGATATTTGGTGGACAGATTCAGATATTTTCAGGCTTGCTTCAGATTC 180
DB 2 GTTCACACAGATATTTGGTGGACAGATTCAGATATTTTCAGGCTTGCTTCAGATTC 180
CY 3 TGGTGATGATGATATATTCATATGATGATGATGATGATGATGATGATGATGATGAT 140
DB 3 TGGTGATGATGATATATTCATATGATGATGATGATGATGATGATGATGATGATGAT 140
CY 4 AGTATATGATGATATTTGATGATATTTGATGATATTTGATGATATTTGATGATATTT 147
DB 4 AGTATATGATGATATTTGATGATATTTGATGATATTTGATGATATTTGATGATATTT 147
CY 5 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
DB 5 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
CY 6 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
DB 6 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
CY 7 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 427
DB 7 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 427
CY 8 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
DB 8 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
CY 9 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 487
DB 9 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 487
CY 10 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
DB 10 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
CY 11 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 547
DB 11 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 547
CY 12 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
DB 12 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
CY 13 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 607
DB 13 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 607
CY 14 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
DB 14 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
CY 15 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 667
DB 15 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 667
CY 16 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
DB 16 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
CY 17 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 727
DB 17 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 727
CY 18 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
DB 18 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
CY 19 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 787
DB 19 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 787
CY 20 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
DB 20 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
CY 21 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 847
DB 21 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 847
CY 22 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
DB 22 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
```

```

848 ATTTTGGATGACATTTTCATCTTTTATGTGTGTTTCTGTAGTTGATCCGAGGAAAA 907
901 GAGTATAGTACCTCGAGAAATCAGAGATGGAGATTTTAGTGTAGGCTTATGATAATTA 960
906 GAGTATAGTACCTCGAGAAATCAGAGATGGAGATTTTAGTGTAGGCTTATGATAATTA 967
941 CCCGCGGTGTGTGTAGAAAAGTATGTAAATTTGCTCTGTTTAAAGACTTGAACATACC 1220
948 CCCGCGGTGTGTGTAGAAAAGTATGTAAATTTGCTCTGTTTAAAGACTTGAACATACC 1227
1001 TCAGAAAGAGAAATTAATAGAAATTTGTAATGTTTCAGAGTATCTCGAAAGAGAAAT 1080
1008 TCAGAAAGAGAAATTAATAGAAATTTGTAATGTTTCAGAGTATCTCGAAAGAGAAAT 1087
1091 TTTTGTAAATAGTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
1098 TTTTGTAAATAGTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1147
1141 TACGATGTTTAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
1148 TACGATGTTTAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1207
1201 TACGATGTTTAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
1208 TACGATGTTTAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1267
1241 CATTCGAGTAATGAATCTG 1241
1268 CATTCGAGTAATGAATCTG 1298

RESULT 2
AF321617 3109 bp mRNA linear PRI 07 DEC-2000
LOCUS AF321617 Homo sapiens unknown mRNA.
DEFINITION AF321617
ACCESSION AF321617
VERSION AF321617.1 GI:11596417
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3109)
Nuyen,M.H. and Liu,C.
FULL-LENGTH cDNA for Unknown#3
Unpublished
REFERENCE 2 (bases 1 to 3109)
Nuyen,M.H. and Liu,C.
DIRECT SUBMISSION
Submitted (15-NOV-2000) Shotgun sequencing, University of
California, Los Angeles, 1541 Le Conte Ave., Los Angeles, CA
90095, CA
Location/Qualifiers:
1..3109
/organism "Homo sapiens"
/mol_type "mRNA"
/db_xref="taxon:9606"
1759..2295
/codon_start 1
/product="unknown"
/protein_id="AAG38612.1"
/db_xref="GI:11596418"
/translation="VASPLGVSGPPGPPGPPPPPPGQASIIQAPGAPRSSSTL
VQEPSEFACFASLVSCVQVGTQDEETRTGVQCQIKPLD:ARQTECFFQKRQCL
SVQPEGVREDVSELRNEVRKQDAVQKRLTKLRHQVQVLESDYVGRKRPDIQGS
LAYLEQASANIAPAKPT"
BASE COUNT 980 a 503 c 621 g 1005 t
ORIGIN

Query Match 100.0% Score 128.; DB 9; Length 3109;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 128.; Conservative 0; Mismatch 0; Gaps 0;

```

```

QY 1 CAACATGCGCGCTCCACTAGGGGTATGTTTCTGGGACACCCACCGGTCTCCCTCAAG 60
DB 1754 CAACATGCGCGCTCCACTAGGGGTATGTTTCTGGGACACCCACCGGTCTCCCTCAGG 1813
QY 61 CCGCGCGCGCTCCCGGCGCAAGCTTCGCTTCTTCAGGAGAGCTCCAGGGCTCTCTACAG 128
DB 1814 CCGCGCGCGCTCCCGGCGCAAGCTTCGCTTCTTCAGGAGAGCTCCAGGGCTCTCTAGAG 1873
QY 121 CTTCAGAGACTTCTTGGTGTGACGAGTTGAGAGTCACTTTTCGAGGGCTTGTTCACATCTC 180
DB 1814 CTTCAGAGACTTCTTGGTGTGACGAGTTGAGAGTCACTTTTCGAGGGCTTGTTCACATCTC 1933
QY 181 TGGTGTGTGAGAGCTATGTCATATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
DB 1914 TGGTGTGTGAGAGCTATGTCATATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1993
QY 241 AGTGTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
DB 1994 AGTGTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2053
QY 254 GATTCAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
DB 2054 GATTCAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2113
QY 361 GATTCAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
DB 2114 GATTCAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2173
QY 421 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
DB 2174 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2233
QY 481 GCTCTTGGCTTACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
DB 2234 GCTCTTGGCTTACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2293
QY 541 GAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
DB 2294 GAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2353
QY 601 TCGTCTTGGCTTACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
DB 2354 TCGTCTTGGCTTACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2413
QY 661 TTTTATGTTCCCATCTGAAAAATTTTCCACTATTTTATAAGCTGTTTAAATTTTGGAGTAC 720
DB 2414 TTTTATGTTCCCATCTGAAAAATTTTCCACTATTTTATAAGCTGTTTAAATTTTGGAGTAC 2473
QY 721 TTTTATGCTGCTGTAGCTTGGATTAACCAAGTAAGTATTTTCTTCTGTTTCTAGCAAA 780
DB 2474 TTTTATGCTGCTGTAGCTTGGATTAACCAAGTAAGTATTTTCTTCTGTTTCTAGCAAA 2533
QY 781 AGTTTACAGCTGTGAATATGATGACAGAGATCTTTTATGGTGGCTTCTGCTGTTTAA 840
DB 2534 AGTTTACAGCTGTGAATATGATGACAGAGATCTTTTATGGTGGCTTCTGCTGTTTAA 2593
QY 841 ATTCTTGCATGACTTTTCTATCTTTTATGTGTGTTTCTGTAGTTTGAATCCGAGGAAAA 900
DB 2594 ATTCTTGCATGACTTTTCTATCTTTTATGTGTGTTTCTGTAGTTTGAATCCGAGGAAAA 2653
QY 901 GAGTATAGTACCTCGAGAAATCAGAGATGGAGATTTTAGTGTAGGCTTATGATAATTA 960
DB 2654 GAGTATAGTACCTCGAGAAATCAGAGATGGAGATTTTAGTGTAGGCTTATGATAATTA 2713
QY 961 CCCGCGGTGTGTGTAGAAAAGTATGTAAATTTGCTCTGTTTAAAGACTTGAACATACC 1020
DB 2714 CCCGCGGTGTGTGTAGAAAAGTATGTAAATTTGCTCTGTTTAAAGACTTGAACATACC 2773
QY 1021 TCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
DB 2774 TCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2833
QY 1081 TTTTGTGAAAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140

```





```
PC C07K16/18,
PC C12N1/21,C12N5/10,C12P21/22,C12Q/68,G01N33/68//A61K38/50, PC
C12N5/50,
PC C12N5/50,A61K37/02
CC Human nucleic acid sequence originating in normal mammary CC
tissue
FH Key Location/Qualifiers
FT source 1..107;
Location/Qualifiers
1..107;
/organism: "Homo sapiens (human)"
/organism: "Homo sapiens"
/mol_type: "genomic DNA"
/db_xref: "taxon:9606"
BASE COUNT 281 a 255 g 317 t
ORIGIN
Query Match 74.9%; Score 1011; DS 6; Length 1071;
Best local Similarity 99.9%; Pred. No. 0;
Matches 1061; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CAACAATGGGGGTCACACTAGAGTGGTATGTTTCTGGGAGCCACCGGGTCCCGCTCAG 60
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5 CAACAATGGGGTCCCACTAGTGGTATGTTTCTGGGAGCCACCGGGTCCCGCTCAG 64
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 CCCCAGCGGCGCTTCCGGGCAAGTTCGTTCTTCAGGCAGTCCAGGCGTCCCTAGAC 120
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 65 CCCCAGCGGCGCTTCCGGGCAAGTTCGTTCTTCAGGCAGTCCAGGCGTCCCTAGAC 124
QY 121 CTTCACAGTACTTCTGGAGCAAGTTCGTTCTTCAGGCAGTCCAGGCGTCCCTAGAC 180
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 125 CTTCACAGTACTTCTGGAGCAAGTTCGTTCTTCAGGCAGTCCAGGCGTCCCTAGAC 184
QY 161 TGGTGAATTCAGAGTATGTAATGACGATCAGGAGGATTCGAGGCGTTCGATC 240
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 165 TGGTGAATTCAGAGTATGTAATGACGATCAGGAGGATTCGAGGCGTTCGATC 244
QY 241 AGTTATCCGAAATTTCTGATATTTGATGATGATGATGATGATGATGATGATGAT 300
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 245 AGTTATCCGAAATTTCTGATATTTGATGATGATGATGATGATGATGATGATGAT 304
QY 301 GATTGAGTATCTCCAGAACAGTACGATGATGATGATGATGATGATGATGATGAT 360
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 305 GATTGAGTATCTCCAGAACAGTACGATGATGATGATGATGATGATGATGATGAT 364
QY 421 GCGAGGAGTGGTGGAGGAGTATGATGATGATGATGATGATGATGATGATGATGAT 480
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 425 GCGAGGAGTGGTGGAGGAGTATGATGATGATGATGATGATGATGATGATGATGAT 484
QY 481 GCTCTTGGCTAGCTGGAGGAGGAGTATGATGATGATGATGATGATGATGATGATGAT 540
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 485 GCTCTTGGCTAGCTGGAGGAGGAGTATGATGATGATGATGATGATGATGATGATGAT 544
QY 541 GAGCAAGGCGAGGAGGAGTGGCTATGATGATGATGATGATGATGATGATGATGATGAT 600
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 545 GAGCAAGGCGAGGAGGAGTGGCTATGATGATGATGATGATGATGATGATGATGATGAT 604
QY 601 TCGTTCCTGGAGTGGAGGAGTGGCTATGATGATGATGATGATGATGATGATGATGATGAT 660
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 605 TCGTTCCTGGAGTGGAGGAGTGGCTATGATGATGATGATGATGATGATGATGATGATGAT 664
QY 661 TTTATGCTCCCATTTGAAATTTTCCAGTATTTTATAGTATTTTATAGTATTTTATAGTAT 720
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 665 TTTATGCTCCCATTTGAAATTTTCCAGTATTTTATAGTATTTTATAGTATTTTATAGTAT 724
QY 721 TTTATAGTATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 725 TTTATAGTATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 784
```

```
QY 781 AGTTAGACTGTGAATATGATGACACAGATTCCTTTTATGGTGGCTTTGCTTGTATTAA 840
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 841 ATTTTCATGACTTTTCATCTTTTATGTTGTTTCTCTGTAGTTTATGATCCGAGGAAAA 900
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 901 GAGTATAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 961 CCCCAGCGGCGCTTCCGGGCAAGTTCGTTCTTCAGGCAGTCCAGGCGTCCCTAGAC 1020
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1021 TCAGAGAGGAGGAGTATGTAATGATGATGATGATGATGATGATGATGATGATGAT 1080
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1081 TCAGAGAGGAGGAGTATGTAATGATGATGATGATGATGATGATGATGATGATGAT 1140
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 5
ACCUIS AX015050 1088 bp DNA linear PAT 07 SEP-2000
DEFINITION Sequence 259 from Patent: WO9553040.
ACCUIS AX015050
VERSION AX015050.1 GI:10041189
KEYWORDS Homo sapiens (human);
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 Schmitt A., Specht C., Dahl E., Hinzmann R., Rosenblat A. and
Pilatovsky C.
Human nucleic acid sequences from ovarian tumour tissue
Patent: WO 9553040-A 259 21-OCT-1999;
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILATSKY CHRISTIAN (DE)
FEATURES
Source
1..1088
/organism: "Homo sapiens"
/mol_type: "genomic DNA"
/db_xref: "taxon:9606"
BASE COUNT 302 a 254 g 316 t
ORIGIN
Query Match 74.9%; Score 1005; DS 6; Length 1088;
Best local Similarity 99.9%; Pred. No. 0;
Matches 1055; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CAACAATGGGGGTCACACTAGAGTGGTATGTTTCTGGGAGCCACCGGGTCCCGCTCAG 60
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 5 CAACAATGGGGTCCCACTAGGAGGATGTTTCTGGGAGCCACCGGGTCCCGCTCAG 64
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 CCCCAGCGGCGCTTCCGGGCAAGTTCGTTCTTCAGGCAGTCCAGGCGTCCCTAGAC 120
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 65 CCCCAGCGGCGCTTCCGGGCAAGTTCGTTCTTCAGGCAGTCCAGGCGTCCCTAGAC 124
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 CTTCACAGTACTTCTGGAGCAAGTTCGTTCTTCAGGCAGTCCAGGCGTCCCTAGAC 180
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 125 CTTCACAGTACTTCTGGAGCAAGTTCGTTCTTCAGGCAGTCCAGGCGTCCCTAGAC 184
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 TGGTGAATTCAGAGTATGTAATGACGATCAGGAGGATTCGAGGCGTTCGATC 240
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 185 TGGTGAATTCAGAGTATGTAATGACGATCAGGAGGATTCGAGGCGTTCGATC 244
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 AGTTATCCGAAATTTCTGATATTTGATGATGATGATGATGATGATGATGATGATGAT 300
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 245 AGTTATCCGAAATTTCTGATATTTGATGATGATGATGATGATGATGATGATGATGAT 304
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 301 GATTGAGTATCTCCAGAACAGTACGATGATGATGATGATGATGATGATGATGATGAT 360
```



QY	362	GAATGATTACAGCGAAGATGCACTAGTCCAGAGCACTTGACAAAGCTGAGGCTG	421
DB	370	GAATGATTACAGCGAAGATGCACTAGTCCAGAGCACTTGACAAAGCTGAGGCTG	429
QY	422	GCAGAGGTGCTGGAGGACATCAAGTGCAGCAAAAGCCGCGACATCCCTCAGG	481
DB	430	GCAGAGGTGCTGGAGGACATCAAGTGCAGCAAAAGCCGCGACATCCCTCAGG	489
QY	442	CTCTTGGCTACCTGGAGGAGGATCTGACACATCTGCGACCTTGAGCGCACTG	541
DB	450	CTCTTGGCTACCTGGAGGAGGATCTGACACATCTGCGACCTTGAGCGCACTG	549
QY	542	AGCAAGGCGCAGAGGCACTGGCTATGATGAGTGGCTGATGCTGAGTGGCACT	601
DB	550	AGCAAGGCGCAGAGGCACTGGCTATGATGAGTGGCTGATGCTGAGTGGCACT	609
QY	602	CTCTGCTGGAGTGCATTTGGAGAGACTCTTCCAGCACTTCCAGCACTGATTT	661
DB	610	CTCTGCTGGAGTGCATTTGGAGAGACTCTTCCAGCACTTCCAGCACTGATTT	669
QY	662	TTATGCTCCCATGAAAATTTTGGCACTATTATTAAGCTGTTACTTTCTGACT	721
DB	670	TTATGCTCCCATGAAAATTTTGGCACTATTATTAAGCTGTTACTTTCTGACT	729
QY	712	TTAACAATCTGTGAGCTTGATTAACAAGTAAGTATTTTCTTTCTTTTACAAA	781
DB	720	TTAACAATCTGTGAGCTTGATTAACAAGTAAGTATTTTCTTTCTTTTACAAA	789
QY	782	GTTTAGATGTGAAATGATGATGATGATGATGATGATGATGATGATGATGAT	841
DB	790	GTTTAGATGTGAAATGATGATGATGATGATGATGATGATGATGATGATGAT	849
QY	842	TTTGTGATGATTTTGTATGATGATGATGATGATGATGATGATGATGATGAT	901
DB	850	TTTGTGATGATTTTGTATGATGATGATGATGATGATGATGATGATGATGAT	909
QY	912	AGTATAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	961
DB	920	AGTATAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	969
QY	972	CCCGGGTGGTGTGAGAAAGTATGTAATTTTCTGTTTAAAGCTGTAAGCTAC	1021
DB	980	CCCGGGTGGTGTGAGAAAGTATGTAATTTTCTGTTTAAAGCTGTAAGCTAC	1029
QY	1022	CAAGAGAGGAGTCTAATACAAATTTCTAAGTTT	1056
DB	1030	CAAGAGAGGAGTCTAATACAAATTTCTAAGTTT	1064
DEFINITION	AC006160 Homo sapiens chromosome 4 clone CC481P14 map: 4p16, complete sequence.		
ACCESSION	AC006160		
VERSION	AC006160.9 GI:5701616		
KEYWORDS	HTS.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	1 (bases 1 to 179310)		
AUTHORS	Stone, N.E., Schmutz, J.J., Shang, J., Pennacchio, L.A., Cox, D.R. and Myers, R.M.		
TITLE	Direct Submission		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 179310)		
AUTHORS	Stone, N.E., Schmutz, J.J., Cox, D.R. and Myers, R.M.		
TITLE	Direct Submission		
JOURNAL	Submitted (08-DEC-1998) Department of Genetics, Stanford Human Genome Center, 825 California Avenue, Palo Alto, CA 94304, USA		
REFERENCE	3 (bases 1 to 179310)		
AUTHORS	Stone, N.E., Schmutz, J.J., Shang, J., Pennacchio, L.A., Cox, D.R. and Myers, R.M.		
TITLE	Direct Submission		
JOURNAL	Submitted (04-AUG-1999) Department of Genetics, Stanford Human Genome Center, 4005 Miranda Avenue, Palo Alto, CA 94304, USA		
COMMENT	On Aug 4, 1999 this sequence version replaced gi:4580748. Estimated Total Number of Errors is 0.1.		
FEATURES	Location/Qualifiers		
SOURCE	Library		
BASE COUNT	50023 a 40057 c 39508 g 49725 t		
ORIGIN	Library		
Query Match	69.3% Score 888; DB 9; Length 179310;		
Best Local Similarity	99.9% Pred. No. G;		
Matches	936; Conservative 0; Mismatches 1; Indels 2; Gaps 0;		
QY	343	AGGATGTGTGAGAACTAAGGAATGAATTACAGCGGAAAGATGACCTAGTCCGAGCACT	402
DB	84315	AGGATGTGTGAGAACTAAGGAATGAATTACAGCGGAAAGATGACCTAGTCCGAGCACT	84236
QY	403	TGACAAAGCTGAGGCACTGSCACAGGTGCTGAGGACATCACCTGAGGACCAAAAG	462
DB	84455	TGACAAAGCTGAGGCACTGSCACAGGTGCTGAGGACATCACCTGAGGACCAAAAG	84196
QY	463	CCGCGACATCCCTCAGGGTCCCTGGCCCTACCTGAGGAGGAGCATCTGCAACATCTG	522
DB	84195	CCGCGACATCCCTCAGGGTCCCTGGCCCTACCTGAGGAGGAGCATCTGCAACATCTG	84116
QY	523	CACCTCTGAAGCCCACTGAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	582
DB	84135	CACCTCTGAAGCCCACTGAGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	84076
QY	583	GTGAGGTGAGCCACACACATCTCTCTCTGAGCTTGACATTTTGGAAAGAACTTTT	642
DB	84075	GTGAGGTGAGCCACACACATCTCTCTCTGAGCTTGACATTTTGGAAAGAACTTTT	84216
QY	643	ATATGAGTTCATTTTAGTCTTTATGCTCCCATGAAAATTTTCCACTATTTTATAAG	702
DB	84015	ATATGAGTTCATTTTAGTCTTTATGCTCCCATGAAAATTTTCCACTATTTTATAAG	83956
QY	703	TGTTAAATTTTGGAGTACTTTTAAACATGCTCTGAGTAAACCAAGTAAGTATTT	762
DB	83955	TGTTAAATTTTGGAGTACTTTTAAACATGCTCTGAGTAAACCAAGTAAGTATTT	83896
QY	763	TTTTTTGCTTTAGCAAAATTTTAGACTGTGAATATGATGACACAGATCTTTTTATG	822
DB	83895	TTTTTTGCTTTAGCAAAATTTTAGACTGTGAATATGATGACACAGATCTTTTTATG	83836
QY	823	TGGCTTGGCTTTTAAATTTTGGATGACTTTTCACTTTTATGCTGTTCTCTGTA	882
DB	83835	TGGCTTGGCTTTTAAATTTTGGATGACTTTTCACTTTTATGCTGTTCTCTGTA	83776
QY	883	GTTTTATCCGAGGAAAGAGTATGATAGCTCAGAAATCAGGAGATGAGGATTTTAGT	942
DB	83775	GTTTTATCCGAGGAAAGAGTATGATAGCTCAGAAATCAGGAGATGAGGATTTTAGT	83716
QY	943	TAGGCTTATGATTAATTTACCCCGGGTGTGTGTAGAAAGTATGTAATTTTCTCTGT	1022

```

Db      837.5 TAGGCTTATGATAATACCCCGGGTGTGTGAGAAAGATATGTAATTTCTGTCT 83656
QY      1003 TTAAGACTTTGAACCTACCTCAAGAAGAGGAATTAATACAAATATTGTAATGTTCCAGA 1062
Db      83656 TTAAGACTTTGAACCTACCTCAAGAAGAGGAATTAATACAAATATTGTAATGTTCCAGA 83596
QY      1063 GCTCTCAGATCAGGATTTTGTGTAATAGTTCAGAGAGGATGAGACTGTCTGGGTT 1122
Db      83595 GCTCTCAGATCAGGATTTTGTGTAATAGTTCAGAGAGGATGAGACTGTCTGGGTT 83536
QY      1123 AGTATAGTAATCTTACAGTAGATCCCTAGGTCATGCTGACCTCTGTTGGGATGTT 1182
Db      83535 AGTATAGTAATCTTACAGTAGATCCCTAGGTCATGCTGACCTCTGTTGGGATGTT 83476
QY      1183 TATATTTATGTTGGTGTCTACTTTTTCATATAAGATATAGTGGAGAGTG 1242
Db      83476 TATATTTATGTTGGTGTCTACTTTTTCATATAAGATATAGTGGAGAGTG 83416
QY      1243 ATACGCTAACATTCATACATCTGACGATGATGATG 1281
Db      83415 ATACGCTAACATTCATACATCTGACGATGATGATG 83377

RESULT 8
AX014817 882 bp DNA l1thead PAT 07-SEP-2000
DEFINITION Sequence 3 from Patent WO9953040.
ACCESSION AX014817
VERSION AX014817.1 GI:1004:054
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1.
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinemann,H., Rosenthal,A. and Pilsarsky,C.
TITLE Human nucleic acid sequences from ovarian tumour tissue
JOURNAL Patent: WO 9953040-A 3 21-OCT-1999; SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN HERNO (DE); ROSENTHAL ANDRE (DE); METZEN GES FUER GENOMFORSCHUNG (DE); PILASKY CHRISTIAN (DE)
FEATURES
    source      Location/Qualifiers
    1..882
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
BASE COUNT 238 a 290 c 206 g 248 t
ORIGIN
Query Match 66.9%; Score 857; DS 6; Length: 882;
Best Locu. Similarity 100.0%; Pred. No. 0;
Matches 857; Conservative C; Mismatches 0; Indels 0; Gaps 0;
QY      1 CAACATGCGGCTCCATAGGGGTATGTTTCTGGACGACCGGTCGCCCTCAGG 60
Db      5 CAATATGCGGCTCCATAGGGGTATGTTTCTGGACGACCGGTCGCCCTCAGG 64
QY      61 CCGCGCGGCGCTCCGCGGCAAGCTTCGCTTCTCAGGAGGCTCAGGCGCTCCTAGAC 120
Db      65 CCGCGCGGCGCTCCGCGGCAAGCTTCGCTTCTCAGGAGGCTCAGGCGCTCCTAGAC 124
QY      121 CTCACAGAGTACTTGGTGGACGAGTGGAGCTATCTTCGAGGCTTGCTTCATCTC 180
Db      125 CTCACAGAGTACTTGGTGGACGAGTGGAGCTATCTTCGAGGCTTGCTTCATCTC 184
QY      181 TGGTAGTCAGGACTATGTCAATGSCACCGATCAGGAGAAATTCGAACCGGTTGTC 240
Db      185 TGGTAGTCAGGACTATGTCAATGSCACCGATCAGGAGAAATTCGAACCGGTTGTC 244
QY      241 AGTGATCCAGAGTTCTCGGATATTGCAAGACACAGCAATGTTTTTCTTACAAAAA 300

```

```

Db      245 AGTGATCCAGAAATTTCTCGATATTGCAAGACACAGACAGAAAGTTTCTTACAAAAA 304
QY      301 GATTCGAGTTATTTTCCAGAACCCAGCAATTTATCAAGAGGATGTGTCAAGACTAA 360
Db      305 GATTCGAGTTATTTTCCAGAACCCAGCAATTTATCAAGAGGATGTGTCAAGACTAA 364
QY      361 GGAATGAATACACGCAAGATGACACTATTCAGAGACACTTGACAAAGCTGAGGCTT 420
Db      365 GGAATGAATACACGCAAGATGACACTATTCAGAGACACTTGACAAAGCTGAGGCTT 424
QY      421 GGCACAGTTCCTGTCAGACATCAAGCTGCACACACAAAGACCCGCCACATCCCTCAG 480
Db      425 GGCACAGTTCCTGTCAGACATCAAGCTGCACACACAAAGACCCGCCACATCCCTCAG 484
QY      481 GCTCTCTGTCATTTGATGAGGCAATTTGCAACATCCCTGACACTCTGAAGCGGAGTT 540
Db      485 GCTCTCTGTCATTTGATGAGGCAATTTGCAACATCCCTGACACTCTGAAGCGGAGTT 544
QY      541 GATTAAGATTTGATTAAGTTTGTATTAAGTTGATTAAGTTGATTAAGTTGATTAAG 600
Db      545 GATTAAGATTTGATTAAGTTTGTATTAAGTTGATTAAGTTGATTAAGTTGATTAAG 604
QY      601 TCTCTCTGTCATTTGATGAGGCAATTTGCAACATCCCTGACACTCTGAAGCGGAGTT 660
Db      605 TCTCTCTGTCATTTGATGAGGCAATTTGCAACATCCCTGACACTCTGAAGCGGAGTT 664
QY      661 TTTATGCTGTCATTTGATTAAGTTTGTATTAAGTTGATTAAGTTGATTAAGTTGATTAAG 720
Db      665 TTTATGCTGTCATTTGATTAAGTTTGTATTAAGTTGATTAAGTTGATTAAGTTGATTAAG 724
QY      721 TTTATGATGTTTGTAGTTGATTAAGTTGATTAAGTTGATTAAGTTGATTAAGTTGATTAAG 780
Db      725 TTTATGATGTTTGTAGTTGATTAAGTTGATTAAGTTGATTAAGTTGATTAAGTTGATTAAG 784
QY      781 AGTTATGATGTTTGTAGTTGATTAAGTTGATTAAGTTGATTAAGTTGATTAAGTTGATTAAG 840
Db      785 AGTTATGATGTTTGTAGTTGATTAAGTTGATTAAGTTGATTAAGTTGATTAAGTTGATTAAG 844
QY      841 ATTATTAAGTATGATTTT 857
Db      845 ATTATTAAGTATGATTTT 861

RESULT 9
BX015361 822 bp DNA linear PAT 27-MAR-2000
LOCUS Human proteins and cDNA[4]
DEFINITION BX015361
ACCESSION BX015361
VERSION BX015361.1 GI:22556459
KEYWORDS JP 201218584 A/77.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1.
AUTHORS Kato,S. and Sasaki,M.
TITLE Human proteins and cDNA[4]
JOURNAL Patent: JP 201218584-A 7 14 AUG 2001; CAPAN SCIENCE AND TECHNOLOGY DEPT
COMMENT CS Homo sapiens (human)
PN JP 201218584-A/7
PD 14-AUG-2001
PF 38-FEB-2000 JP 20000311662
PI SEIJI KATO,XIAOYU SAKI
PC C12N15/89,C07K14/435,C07K16/18,C12N1/15,C12N1/19,C12N1/21,PC
C12N5/1577
PC C12N25/28,C12N15/00,C12N5/00
CC CC
FH Key Location/Qualifiers
FT CDS Location/Qualifiers
    1..822
    /organism="Homo sapiens"

```



```

RESULT 11
AF317680      789 bp  mRNA  linear  PRI 21-FEB-2001
LOCUS        Homo sapiens tumor-related protein mRNA, complete cds.
DEFINITION
ACCESSION   AF317680
VERSION     AF317680.1 GI:13021943
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Mammalia; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catartini; Hominoidea; Homo.
REFERENCE   1 (bases 1 to 789)
  AUTHORS   Mail,N. and Liu,C.
  TITLE     A gene from tumor
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 789)
  AUTHORS   Mail,N. and Liu,C.
  TITLE     Cloned cDNA
  JOURNAL   Submitted 01 NOV-2000; Surg. Oncology, University of
  California, Los Angeles, 10833 Le Conte Ave., Los Angeles, CA
  90095, USA
FEATURES             Location/Qualifiers
     source          1..789
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
     CDS             12..593
                     /codon_start=1
                     /product="tumor-related protein"
                     /protein_id="AAK11561.1"
                     /db_xref="GI:13021830"
                     /translations="MAAPLGMSGQPPGHPAPPQFPGACSLQMAPQAPFSSSTL
                     VQLESEFFACFASISQDVNLTQDEEISTGVQDQIQKFLDIASQTFQFFQKRAQL
                     SVQKPEVKEDESEKLNELKDALVQKILTKRHWQVLEIDNVQHKRADIQGS
                     LAYCAGICQHPITSEANVSKQRLAVENAVA"
BASE COUNT        216 a 179 c 194 g 196 t 2 others
ORIGIN
Query Match      38.7%; Score 496; DB 9; Length 789;
Best Local Similarity 100.0%; Pred. No. 5.4; 281;
Matches 496; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AACATGGCGCTCCACCTAGGAGGTATGTTTCTGSGCAGGACCCCTCCCTCAGGCG 67
DB 3 AACATGGCGCTCCACCTAGGAGGTATGTTTCTGSGCAGGACCCCTCCCTCAGGCG 68

QY 63 CCGCGCGGCGCTCCGCGGCCAAGCTTCCTCTTTCCAGGACGCTCCAGCGCTCTTAGACCT 122
DB 63 CCGCGCGGCGCTCCGCGGCCAAGCTTCCTCTTTCCAGGACGCTCCAGCGCTCTTAGACCT 126

QY 123 TCCAGCAGTATTTGGTGGAGCTTGAGTATCTTTCCAGGAGCTTCCTGATCTCTG 182
DB 123 TCCAGCAGTATTTGGTGGAGCTTGAGTATCTTTCCAGGAGCTTCCTGATCTCTG 186

QY 129 TCCAGCAGTATTTGGTGGAGCTTGAGTATCTTTCCAGGAGCTTCCTGATCTCTG 249
DB 129 TCCAGCAGTATTTGGTGGAGCTTGAGTATCTTTCCAGGAGCTTCCTGATCTCTG 249

QY 183 GTGAGTCAGGACTATGTCATGCGACCGGATCAGAGAGAAATTCAGAGCTGTGTGATCAG 242
DB 183 GTGAGTCAGGACTATGTCATGCGACCGGATCAGAGAGAAATTCAGAGCTGTGTGATCAG 249

QY 243 TGTATCCAGAGTTTCTGGATATTGGAAGACAGACAGATCTTTTCTTCTTACAAAAA 302
DB 243 TGTATCCAGAGTTTCTGGATATTGGAAGACAGACAGATCTTTTCTTCTTACAAAAA 308

QY 249 TGTATCCAGAGTTTCTGGATATTGGAAGACAGACAGATCTTTTCTTCTTACAAAAA 308
DB 249 TGTATCCAGAGTTTCTGGATATTGGAAGACAGACAGATCTTTTCTTCTTACAAAAA 316

QY 303 TTGCAGTATCTGTCCCAACACAGACAGATTCAGAGGATGTGTCAGAGACTAAG 362
DB 303 TTGCAGTATCTGTGTCCCAACACAGACAGATTCAGAGGATGTGTCAGAGACTAAG 369

QY 363 AATGAATTACAGCGGAAGATGCACTAGTCAGAGACATTTTCAAGGCTGAGGACTG 422
DB 363 AATGAATTACAGCGGAAGATGCACTAGTCAGAGACATTTTCAAGGCTGAGGACTG 428

QY 423 CAGAGGTTGCTGGAGGACATCAAGCTGCAGCACAAGAGCCGCCACATCTCCCTCAGG 482
DB 423 CAGAGGTTGCTGGAGGACATCAAGCTGCAGCACAAGAGCCGCCACATCTCCCTCAGG 488

```

```

QY 443 TCTCTTCCTCACTTGG 498
DB 443 TCTCTTCCTCACTTGG 504

RESULT 12
AF317678      905 bp  mRNA  linear  PRI 21-FEB-2001
LOCUS        Homo sapiens putative angiogenesis related mRNA sequence.
DEFINITION
ACCESSION   AF317678
VERSION     AF317678.1 GI:13021843
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Mammalia; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catartini; Hominoidea; Homo.
REFERENCE   1 (bases 1 to 905)
  AUTHORS   Mail,N. and Liu,C.
  TITLE     Direct Submission
  JOURNAL   Submitted 01 NOV-2000; Surg. Oncology, University of
  California, Los Angeles, 10833 Le Conte Ave., Los Angeles, CA
  90095, USA
FEATURES             Location/Qualifiers
     source          1..905
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
     gene            1..905
                     /gene="putative angiogenesis-related"
BASE COUNT        204 a 184 c 210 g 111 t 2 others
ORIGIN
Query Match      41.8%; Score 494; DB 9; Length 905;
Best Local Similarity 49.4%; Pred. No. 1.6; 226;
Matches 424; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 CCATAGAGGAGATCTTCTTGGAGAGGAGATCGTCCCTCCAGCGCGCGAGTCT 74
DB 15 CCATAGAGGAGATCTTCTTGGAGAGGAGATCGTCCCTCCAGCGCGCGAGTCT 75

QY 75 CCGGCGAAGCTTCCTCTTCTTCCAGGAGCTTCAGAGGCTCCAGACTCCAGCATAT 114
DB 75 CCGGCGAAGCTTCCTCTTCTTCCAGGAGCTTCAGAGGCTCCAGACTCCAGCATAT 118

QY 115 TTCTGCGAGGCTTGAGTATCTTTCTGAGGATGTTGTTGCTCTCTGGTGAATCAGAG 174
DB 115 TTCTGCGAGGCTTGAGTATCTTTCTGAGGATGTTGTTGCTCTCTGGTGAATCAGAG 178

QY 129 TGTGAGTGAATTCATGATCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 184
DB 129 TGTGAGTGAATTCATGATCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 188

QY 195 TATCTGAATGAGCATATGAGGAGAGATTCAGAGGCTTCAGTATCTATCAGATTCAG 254
DB 195 TATCTGAATGAGCATATGAGGAGAGATTCAGAGGCTTCAGTATCTATCAGATTCAG 258

QY 259 TTTCTGATATTCAGAGAGAGATTCAGAGGCTTCAGTATCTATCAGATTCAGATTCAG 319
DB 259 TTTCTGATATTCAGAGAGAGATTCAGAGGCTTCAGTATCTATCAGATTCAGATTCAG 323

QY 319 TTTCTGATATTCAGAGAGAGATTCAGAGGCTTCAGTATCTATCAGATTCAGATTCAG 378
DB 319 TTTCTGATATTCAGAGAGAGATTCAGAGGCTTCAGTATCTATCAGATTCAGATTCAG 382

QY 374 GCGAGAGATTCAGTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 433
DB 374 GCGAGAGATTCAGTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 438

QY 434 GCGAGAGATTCAGTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 493
DB 434 GCGAGAGATTCAGTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 498

```







Collymore, A., Cook, A., Cooke, P., Corum, B., DeBello, K., Diaz, S., Dodge, K., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, B., Galagan, J., Gardy, S., Gruber, L., Grand, P., Hafez, N., Hager, D., Hager, B., Hall, J., Horton, B., Hume, W., Iliev, I., Johnson, K., Jones, C., Karat, A., Karatas, A., Kellis, C., Landers, J., Levine, P., Lindblad, T., Liu, G., Liu, A., Mabbitt, R., Maclean, C., Macdonald, F., Maier, J., Manning, J., Matthews, S., McCarthy, M., McDermid, J., Menzies, D., Milova, T., Miska, V., Murphy, T., Naylor, J., Nguyen, C., Oliver, J., Orban, C., O'Connor, T., O'Donnell, F., Ornel, F., Peterson, K., Phunkhang, P., Pflieger, N., Rachupka, A., Rasmussen, J., Raymond, C., Retta, A., Rice, C., Rogov, P., Roman, J., Schauer, S., Schuppach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange, Thumann, N., Stojanovic, N., Stubbs, M., Tamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, N., Vassiliev, H., Venkataratnam, S., Vieri, R., Viora, A., Wilson, R., Wu, X., Wyman, D., Young, S., Zainuddin, J., Zerbek, L., Zimmer, A., and Zody, M.

Center: Whitehead Institute for Genome Research  
 Center code: WIR  
 Web site: <http://www.seq.wi.mit.edu>  
 Contact: [sequence.submission@genome.wi.mit.edu](mailto:sequence.submission@genome.wi.mit.edu)  
 Project Information  
 Center Project name: 13407  
 Center Clone name: 7 D 3  
 Summary Statistics  
 Sequencing vector: Plasmid; n/a; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.96231  
 Consensus quality: 214508 bases at least C40  
 Consensus quality: 215611 bases at least C30  
 Consensus quality: 216108 bases at least Q20  
 Insert size: 134000; average 134000  
 Insert size: 216592; sum of contigs  
 Quality coverage: 13.0 in Q30 bases; average 13.0  
 Quality coverage: 11.0 in Q30 bases; sum of contigs

# TITLE JOURNAL

## COMMENT

All repeats were identified using RepeatMasker:

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

FEATURES  
 source  
 1. 217692  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /clone="RP23-7D3"  
 /clone\_id="RP23-7D3" Female Mouse BAC"  
 1. 10343  
 /note="assembly fragment"  
 clone\_end:SP6  
 vector\_side:left"  
 20144..11973  
 /note="assembly fragment"  
 12773..14204  
 /note="assembly fragment"  
 14305..120815  
 /note="assembly fragment"  
 20916..73431  
 /note="assembly fragment"  
 73532..102829  
 /note="assembly fragment"  
 102930..119343  
 /note="assembly fragment"  
 119344..114300  
 /note="assembly fragment"  
 134401..155072  
 /note="assembly fragment"  
 155173..183047  
 /note="assembly fragment"  
 183148..214210  
 /note="assembly fragment"  
 214311..217692  
 /note="assembly fragment"  
 clone\_end:T7  
 vector\_side:right"  
 BASE COUNT 6358 a 46947 c 46159 g 59728 t 1103 others  
 ORIGIN

Query Match 4.43; Score 56; DB 2; Length 217692;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-20;  
 Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QV 190 GTCCAGGAGGAGCTTACAAAGCTGAGCGCATTCGACAGCAGTGTGTCGAGGACATCAA 445  
 Db 141199 GTCCAGGAGGAGCTTACAAAGCTGAGCGCATTCGACAGCAGTGTGTCGAGGACATCAA 341254

RESULT 17  
 AC005063  
 LOCUS  
 DEFINITION Homo sapiens chromosome UNK clone CTB-99B5, WORKING DRAFT SEQUENCE.  
 6 unordered pieces.  
 AC005063  
 AC005063  
 AC005063.2 GI:13431268  
 HTO: HTO; PHASE: HTO; DRAFT: HTO; ACTIVE: HTO.  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 1: bases 1 to 127188;  
 Waterston, R.H.  
 The sequence of Homo sapiens clone  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 127188)  
 Waterston, R.H.  
 DIRECT SUBMISSION  
 Submitted (12-JUN-1998) Genore Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 COMMENT On Mar 22, 2001 this sequence version replaced gi:1212928.

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H RG099505
----- Summary Statistics -----
Sequencing vector: M13: 981
Sequencing vector: Plasmid: 21
Chemistry: Dye-terminator S1: 72% of reads
Assembly program: Phrap: version 0.990319
Consensus quality: 124785 bases at least Q40
Consensus quality: 125506 bases at least Q30
Consensus quality: 125754 bases at least Q20
Insert size: 0000; agarose-1p
Insert size: 126106; sum-of-coverage
Quality coverage: in 320 bases; agarose-1p
Quality coverage: in 320 bases; sum of coverage
-----
* NOTE: This is a "working draft" sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 4050: contig of 4050 bp in length
* 4150: gap of unknown length
* 4151: contig of 5664 bp in length
* 9914: gap of unknown length
* 9915: contig of 18425 bp in length
* 28339: contig of 24401 bp in length
* 28439: gap of unknown length
* 28440: contig of 24401 bp in length
* 52840: gap of unknown length
* 52841: contig of 73566 bp in length
* 52941: 126506: gap of unknown length
* 126507: 127188: contig of 532 bp in length.
*
* Location/Qualifiers
*   1: 127188
*     /organism="Homo sapiens"
*     /mol_type="genomic DNA"
*     /db_xref="taxon:9606"
*     /contosome="UNK"
*     /clone="CTS-99B5"
*   .. 4050
*     /note="assembly_name:Contig25"
*   4151: 9914
*     /note="assembly_name:Contig26"
*   9915: 28339
*     /note="assembly_name:Contig27"
*   28440: 52840
*     /note="assembly_name:Contig28"
*   52941: 126506
*     /note="assembly_name:Contig29"
*     clone_end: 3p6
*     vector_side: left
*   126507: 127188
*     /note="assembly_name:Contig36"
*
* BASE COUNT 36807 a 25531 c 26056 g 38248 t 546 others
*
* ORIGIN
*
* Query Match 3.11% Score 40; DB 2; Length 127188;
* Best Local Similarity 100.0%; Pred. No. 5.4e-11;
* Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 839 AAATTTTCATGACCTTTTCATCTTTTATGCTGTTCC 878
|||||
DB 42977 AAATTTTCATGACCTTTTCATCTTTTATGCTGTTCC 43016
|||||
RESULT 18

```

```

AL359844
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL

```

## COMMENT

AL359844 Human DNA sequence from Clone RP11-31418 on chromosome 10, complete sequence.

AL359844 AL359844.15 G11131840B3

HTG

Homo sapiens (human)

Homo sapiens

Exkaryota; Metazoa; Chordata; Clamata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 171364)

Almeida J.

Direct Submission

Submitted (1 DEC 2001) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB9 9SA, UK. E-mail enquiries: [clonerequests@sanger.ac.uk](mailto:clonerequests@sanger.ac.uk)

On Dec 5, 2001 this sequence version replaced g111761332.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality > 30). An attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one phosid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following annotations are used to associate primary accession numbers given in the feature table with their source databases: EMBL, EXFBI, SWISSPROT, TrEMBL, WPI, WAXPDB; information on the WAXPDB database can be found at <http://www.sanger.ac.uk/Projects/Celegans/wormpep/>

This sequence was generated from part of a bacterial clone library of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Ch10>

RP11-31418 is from the library RP01-11.2 constructed by the group of Peter de Jong. For further details see <http://www.chori.org/bacpac/homo.htm>

Vector: pNAc16

This sequence is the entire insert of clone RP11-31418. The left end of clone RP11-259516 is at 111643 in this sequence. The true right end of clone RP11-437418 is at 13148 in this sequence.

## FEATURES

## Source

```

1: 171364
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/contosome="10"
/clone="RP11-31418"
/clone_id="RP01-11.2"
/clone_lib="RP01-11.2"

```

```

BASE COUNT 47446 a 44344 c 36983 g 52361 t
ORIGIN

```

```

Query Match 3.11% Score 40; DB 2; Length 171364;
Best Local Similarity 100.0%; Pred. No. 5.4e-11;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 839 AAATTTTCATGACCTTTTCATCTTTTATGCTGTTCC 878
|||||
DB 133862 AAATTTTCATGACCTTTTCATCTTTTATGCTGTTCC 13390:

```

```

RESULT 19
HS22715/c
LOCUS
DEFINITION
ACCESSION

```

HS22715 Human DNA sequence from Clone RP6-22715 on chromosome Xp11.22-11.3, complete sequence.

AL031585







```

repeat_region 6481..6579
/note="L2 repeat: matches 2634..2743 of consensus"
repeat_region 6626..7016
/note="LTR repeat: matches 6..390 of consensus"
repeat_region 7655..7652
/note="44 copies 2 mer ct 73% conserved"
repeat_region 7668..7781
/note="57 copies 2 mer ct 75% conserved"
misc_feature complement(7844..8286)
/note="Tandem repeat region: contains forced join and single clone. HindIII digest suggests there is approximately 1kb missing from the tandem repeat."
repeat_region 8095..8132
/note="19 copies 2 mer ca 78% conserved"
misc_feature complement(8206..8578)
/note="match: GSS: Em:AO535283"
misc_feature complement(8211..8572)
/note="match: GSS: Em:AO537005"
misc_feature complement(8265..8579)
/note="match: GSS: Em:AO487567"
misc_feature 8822..9424
/note="CPG island"
/note="evidence=not_experimental"
repeat_region 10072..10203
/note="FAM_C repeat: matches 2..133 of consensus"
repeat_region 10737..11028
/note="AluX repeat: matches 6..298 of consensus"
repeat_region 11364..11671
/note="AluX repeat: matches 1..104 of consensus"
repeat_region 12208..12496
/note="AluX repeat: matches 1..299 of consensus"
repeat_region 12508..12811
/note="AluX repeat: matches 1..301 of consensus"
repeat_region 13203..13290
/note="LTR repeat: matches 243..317 of consensus"
repeat_region 13319..13366
/note="LTR repeat: matches 230..327 of consensus"
repeat_region 13394..13480
/note="LTR repeat: matches 468..547 of consensus"
repeat_region 13501..13636
/note="FLAV_C repeat: matches 1..133 of consensus"
repeat_region 13828..14026
/note="AluX repeat: matches 1..304 of consensus"
misc_feature complement(13929..14310)
/note="match: GSS: Em:AO756915"
repeat_region 14518..14584
/note="L2 repeat: matches 2677..2745 of consensus"
repeat_region 14591..14757
/note="L2 repeat: matches 1709..1875 of consensus"
repeat_region 14937..15005
/note="AluJ/FLAM repeat: matches 1..73 of consensus"
misc_feature complement(15027..15331)
/note="match: GSS: Em:AO280927"
repeat_region 15040..15332
/note="AluX repeat: matches 1..293 of consensus"
misc_feature complement(15040..15332)
/note="match: STS: Em:H552819S"
repeat_region 15067..15333
/note="match: GSS: Em:A2518424"
misc_feature complement(15065..15333)
/note="match: GSS: Em:AO446595"
misc_feature complement(15066..15346)
/note="match: GSS: Em:AO807616"
misc_feature complement(15089..15356)
/note="match: GSS: Em:AO490556"
misc_feature 15091..15333
/note="match: STS: Em:G20204"
misc_feature complement(15097..15332)
/note="match: GSS: Em:A2521752"
misc_feature 15109..15332
/note="match: GSS: Em:B01680"
misc_feature 15111..15332
/note="match: GSS: Em:A2521733"

```

```

misc_feature 15114..15332
/note="match: GSS: Em:AO52998"
misc_feature 15117..15332
/note="match: GSS: Em:A2519263"
misc_feature 15121..15332
/note="match: GSS: Em:AO223462 Em:AO631465"
misc_feature 15121..15331
/note="match: GSS: Em:A2520451"
misc_feature complement(15127..15380)
/note="match: GSS: Em:AO15261"
misc_feature 15129..15340
/note="match: GSS: Em:AO532262"
misc_feature complement(15135..15332)
/note="match: GSS: Em:AO616206"
misc_feature 15143..15329
/note="match: GSS: Em:AO591113"
misc_feature 15154..15348
/note="match: GSS: Em:A1335485"
misc_feature complement(15161..15332)
/note="match: STS: Em:AO4454"
misc_feature complement(15171..15347)
/note="match: GSS: Em:AO44808"
misc_feature complement(15171..15343)
/note="match: STS: Em:G55778"
misc_feature complement(15182..15328)
/note="match: GSS: Em:AO15996"
misc_feature complement(15198..15377)
/note="match: STS: Em:AO57766"
misc_feature 15199..15332
/note="match: GSS: Em:AO279650"
misc_feature complement(15194..15356)
/note="match: GSS: Em:AO281971"
misc_feature complement(15200..15331)
/note="match: GSS: Em:AO88705"
misc_feature 15207..15381
/note="match: STS: Em:H5196VD10"
misc_feature 15210..15353

```

Query Match: 2.51; Score 37; GB 3; Length 122867;

Best Local Similarity 100.0%; Prog. No. 1.0e-06;

Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

CY 177 TATCTGTCATGTCAGACATGTCATGTCAC 208
      ||||| ||||| ||||| ||||| |||||
CB 76834 TATCTGTCATGTCAGACATGTCATGTCAC 76833

```

# RESULT 23

```

AP000725/c
LOCUS      108444 bp    DNA    linear    PRI 15 MAR 2003
DEFINITION Homo sapiens genomic DNA, chromosome 11q, clone:PM9-2012, complete
            sequence.
ACCESSION  AP000725
VERSION     AP000725.4  GI:14519568
KEYWORDS    HTG
SOURCE      Homo sapiens
            Human
ORGANISM    Homo sapiens
            Chordata; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
            Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Soog,P.,
            Fujiyama,A., Yada,T., Tetsuki,Y., Watanabe,H. and Sakaki,Y.
            Homo sapiens genomic DNA
            Published only in Database (1999).
REFERENCE   2
            Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Soog,P.,
            Fujiyama,A., Yada,T., Tetsuki,Y., Watanabe,H. and Sakaki,Y.
            Direct Submission
            Submitted (16-NOV-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
            1-7-22 Sueno-cho,Tsukumi-ks, Yokohama, Kanagawa 230-0445, Japan
            E-mail:hattori@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170;

```

COMMENT On Jun 20, 2001 this sequence version replaced g1:11136463.  
 FEATURES location/Qualifiers  
 1..108444  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="11"  
 /map="11q"  
 /clone="CM99-2013"  
 BASE COUNT 3559 A 1950 C 19828 G 4519 T  
 ORIGIN  
 Query Match 2337 Score 29; EB 97; Length 108444  
 Best local similarity 100.0%; Pref.No. 0.00077;  
 Matches 29; Conservative 0; Labels 0; Gaps 0;  
 413 CTGTCAGAAACCAAGCAAGTATATCAA 41  
 ||||| ||||| ||||| ||||| |||||  
 17973 CTGTCAGAAACCAAGCAAGTATATCAA 97945

RESULT 24  
 ACC17071  
 DEFINITION Homo sapiens BAC clone RP11-410E4 from 2, complete sequence.  
 ACCESSION ACC17071  
 VERSION ACC17071.9 GI:14589725  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Fukuyama; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo;  
 1 (bases 1 to 143456)  
 AUTHORS Rubston, J.E. and Waterston, R.  
 TITLE Toward a complete human genome sequence  
 JOURNAL Genome Res. 8 (11): 109-1106 (1998)  
 MEDLINE 99063792  
 PUBMED 9847074  
 REFERENCE 2 (bases 1 to 143456)  
 AUTHORS Vandrunft, A. and Cotton, M.  
 TITLE The sequence of Homo sapiens BAC clone RP11-410E4  
 JOURNAL Unpublished (2001)  
 REFERENCE 3 (bases 1 to 143456)  
 AUTHORS Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-DEC-1999) Genome Sequencing Center, Washington University School of Medicine, 444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE 4 (bases 1 to 143456)  
 AUTHORS Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (03-JUL-2001) Genome Sequencing Center, Washington University School of Medicine, 444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE 5 (bases 1 to 143456)  
 AUTHORS Waterston, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington University, 444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 REFERENCE 6 (bases 1 to 143456)  
 AUTHORS Waterston, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-MAR-2002) Department of Genetics, Washington University, 444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 COMMENT On Jul 3, 2001 this sequence version replaced g1:13518260.  
 ----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc>  
 Contact: sapiens@wustl.wustl.edu  
 ----- Summary Statistics  
 -----  
 Center project name: H\_NH0410504

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats, all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RP11-410E4 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatem, M., Catalanese, J.J., and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-9. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Peter de Jong and coworkers at <http://www.chori.org>  
 VECTOR: pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP11-66C4. Actual start of this clone is at base position 1 of RP11-410E4; actual end is at base position 16339 of RP11-66C4.

H\_NH0410504 from base positions 103516 to 103641 is represented by sequence derived from PCR.

#### FEATURES

location/Qualifiers  
 1..163456  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 /map="2"  
 /clone="RP11-410E4"  
 /clone\_idb="RP11-410E4"  
 1..1688  
 /rpt\_family="ERV1"  
 723..1997  
 /rpt\_family="MER1\_type"  
 1094..1212  
 /rpt\_family="MER1\_type"  
 1213..1786  
 /rpt\_family="ERV1"  
 1743..12143  
 /note="match to EST AW900441 (NID:98064646)"  
 1787..1866  
 /rpt\_family="MER1\_type"  
 2493..2651  
 /rpt\_family="ERV1"  
 2740..3046  
 /rpt\_family="Alu"  
 3187..3285  
 /note="similar to EST AL529206 (NID:g12792699)"  
 12326..12456  
 /rpt\_family="A-rich"  
 12446..13451  
 /rpt\_family="L1"  
 13452..14278  
 /rpt\_family="L1"  
 14264..14290  
 /rpt\_family="AT-rich"  
 14293..14291









Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrum, J., Meneses, J., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C., O'Connor, T., O'Neill, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, K., Raymond, C., Rettar, R., Riese, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, P., Seaman, S., Sever, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodor, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, G., Zembek, I., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (28-NOV-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 167446)

REFERENCE  
AUTHORS  
Birken, B., Nusbaum, C., Landers, E., Abuelkheil, A., Allen, N., Anderson, S., Arachchi, H. Y., Barua, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collyer, A., Cook, A., Cooke, P., Cotnam, B., DeAtelliano, K., Diaz, C. S., Dodge, S., Dooley, K., Dorris, E., Erickson, J., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagani, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hages, R., Hall, C., Horton, J., Hulme, W., Iliev, I., Johnson, P., Jones, C., Karat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Liu, G., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrum, J., Meneses, J., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Neill, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, K., Rachevka, A., Ramasamy, U., Raymond, C., Rettar, R., Riese, C., Rogov, P., Roman, J., Schauer, S., Schupback, S., Seaman, S., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodor, J., Topham, K., Travers, M., Vassiliev, H., Venkatacham, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, G., Zembek, I., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (27-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jan 28, 2003 this sequence version replaced g12751764.  
All repeats were identified using RepeatMasker:  
Smith, A. F. A. & Green, P. (1996, 1997)  
<http://ftp.genome.washington.edu/JM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: W-BR  
Web site: <http://www.seql.wi.mit.edu>  
Contact: sequence submissions@genome.wi.mit.edu  
Project information  
Center project name: L4440  
Center clone name: 5.12.0

NOTE: This is a working draft sequence. It currently consists of 8 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.  
This sequence will be replaced  
by the finished sequence as soon as it is available and the accession number will be preserved.

12611: contig of 12610 bp in length  
12711: contig of 3873 bp in length  
16584: contig of 100 bp  
16584: gap of 100 bp  
16584: contig of 6567 bp in length  
23351: contig of 100 bp  
23451: gap of 100 bp  
23451: contig of 21025 bp in length  
44376: gap of 100 bp  
44376: contig of 4580 bp in length  
49156: gap of 100 bp  
49256: contig of 17766 bp in length  
67022: gap of 100 bp  
67122: contig of 60447 bp in length

127569 127568: gap of 100 bp  
127669 167446: contig of 39778 bp in length.

FEATURES  
Source  
1..167446  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/map="11"  
/clone="RP11-5L12"  
/clone\_lib="RP11-5L12"  
/clone\_map="37030 g 50854 c 917 others"

BASE COUNT  
ORIGIN  
439.8 a 34727 c 37030 g 50854 t 917 others

Query Match  
Best Local Similarity 100.0%; Score 23; DB 2; Length 167446;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 763 TTTTCTTTTCTCTTTAGCAAAAG 782  
DQ 156276 TTTTCTTTTCTCTTTAGCAAAAG 156254

RESULT 29  
AC079296/c  
AC079296/c  
LOCUS  
DEFINITION  
Homo sapiens chromosome 11, clone RP11-467K18, complete sequence.  
AC079296  
VERSION  
AC079296.5 G:28261647  
KEYWORDS  
HTG.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 194686)  
Birken, B., Nusbaum, C. and Landers, E.  
Homo sapiens chromosome 11, clone RP11-467K18  
Unpublished  
2 (bases 1 to 194686)  
Birken, B., Linton, L., Nusbaum, C., Lander, E., Abrahams, H., Allen, N., Anderson, S., Barua, N., Bastien, V., Bida, P., Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Chappel, Y., Collange, C., Collins, S., Collymore, A., Cooke, P., DeAtelliano, K., Dewar, K., Diaz, J. S., Dodge, S., Ferreira, P., FitzHugh, W., Gage, D., Galagani, J., Gardyna, S., Girdle, S., Goyette, M., Graham, L., Grand-Pierre, N., Hages, B., Heaford, A., Horton, J., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Lacroque, K., Lamazares, R., Landers, T., Lebecky, J., Levine, R., Lindblad-Toh, K., Macdonald, P., Marquis, N., McCarthy, M., McSwan, P., McKenna, K., McNeeters, R., Meldrum, J., Meneses, J., Mihova, T., Mlenga, V., Norman, C., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Neill, P., O'Neill, D., Pollara, V., Raymond, C., Rieback, M., Riley, R., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Rieback, M., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Sever, P., Sognez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodor, J., Tirrell, A., Travers, M., Trigglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, G., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (26-AUG-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 194686)  
Birken, B., Nusbaum, C., Landers, E., Ali, A., Allen, N., Anderson, S., Barua, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., DeAtelliano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagani, J., Gardyna, S., Girdle, S., Horton, J., Hages, B., Grand-Pierre, N., Hafez, N., Hages, R., Karat, A., Karatas, A., Kells, C., Landers, T., Johnson, R., Jones, C., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrum, J., Meneses, J., Mihova, T.,



```

repat_region      /rpt_family="L2"
20747..21504
repat_family="L1MB"
21305..21309
repat_region      /rpt_family="Alus1"
21310..21803

Query Match
Best Local Similarity 100.00; Freq.No. 0.62; Length 194656;
Matches 23; Conservative 0; Mismatches 0; Inverts 0; Gaps 0;

CY 350 TTTTCTTTTCTTTTATCAAAAG 982
||||| ||||| ||||| |||||
CG 12625 TTTTCTTTTCTTTTATCAAAAG 12625

RESULTS
AC026077/C AC026077 19923 bp DNA linear HTG 07-JUL-2000
DEFINITION Homo sapiens chromosome 11 clone RP11-299K14, WORKING DRAFT
SEQUENCE 29 unordered pieces.
AC026077 AC026077.4 G1:R599094
VERSION HTG; HTGS PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
Xereston R.H.
The sequence of Homo sapiens clone
Submitted (19-MAR-2000) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT
On May 26, 2000 this sequence version replaced g1:774912.

..... Genome Center .....
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/cgi-bin/index.shtml
Genetic project name: H.MIM:60814
..... Human Genome Project .....
Sequencing vector: pUC19
Characteristics: Dye-primed 27, 100% of reads
Assembly program: Phrap; version 0.99.01.9
Consensus quality: 178194 bases at least Q40
Consensus quality: 184169 bases at least Q30
Consensus quality: 187367 bases at least Q20
Insert size: 197500; average=19
Insert size: 196400; sum-of-contrasts
Quality coverage: 4.00 in Q20 bases; maximum=17
Quality coverage: 4.13 in Q20 bases; sum-of-contrasts
.....
NOTE: This is a "Working Draft" sequence. It currently
consists of 29 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
.....
1594: contig of 1594 bp in length
1595 1594: gap of unknown length
1695 1695: contig of 1695 bp in length
3102 3102: gap of unknown length
3402 3402: contig of 1549 bp in length

```

```

FEATURES
            source
1..199203
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chr="osome=11"
    /clone="RP11-299K14"
1..1594
    /note="assembly_name:Contig10"
1695..3101
    /note="assembly_name:Contig11"
3402..5350
    /note="assembly_name:Contig12"
5451..7179
    /note="assembly_name:Contig13"
7280..8571
    /note="assembly_name:Contig14"
8674..10055
    /note="assembly_name:Contig15"
clone_end:SP6
vector_start:right

```



contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
Project Information  
Center project name: 3896  
Center clone name: CH210-4P6  
Summary Statistics

Assembly program: Atlas  
Consensus quality: 236074 bases at least 90  
Consensus quality: 237160 bases at least 90  
Consensus quality: 238144 bases at least 90  
Estimated insert size: 44449; sum of contigs estimation  
Quality coverage: 1x in 320 bases; sum of contigs estimation

NOTE: Estimated insert size may differ from sequence length (see [http://www.hgsc.bcm.tmc.edu/databases/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/databases/genbank_draft_data.html)).  
NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 242228: contig of 242028 bp in length  
2 242129 242128: gap of unknown length  
3 242129 243449: contig of 1121 bp in length

## FEATURES

source

location/Qualifiers  
1..243249  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10106"  
/clone="CH210-5B6"

1..1524  
/note="wgs end extension  
clone end: 7"  
6549..7345  
/note="clone boundary  
clone end: 7"  
site:score1

BASE COUNT 55398 1 42368 0 41574 1 1004 - 1000 10000  
ORIGIN

Query Match 1.84; Score 101; E=2; Ident 100.00  
Best Local Similarity 100.00; Prev. Max. Ident 100.00  
Matches 23; Conservative 0; Mismatches 0; Invert 0; Gaps 0;

QY 474 CTCAGGCGCTCTTGGCTACT 496  
|||||  
DB CTCAGGCGCTCTTGGCTACT 496

RESULT 12  
AC111223  
LOCUS  
DEFINITION  
AC111223 242190 bp DNA linear HTG 13-MAY-2003  
Pattus norvegicus clone CH210-12065, \*\*\* SEQUENCING IN PROGRESS  
\*\*\* 14 unordered pieces.

AC111223  
AC111223.5 GI:30578497  
HTG; HTGS, PHASE1; HTGS, DRAFT; HTGS, ENRICHED.  
Pattus norvegicus (Norway rat)  
Pattus norvegicus  
ORGANISM  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE  
1 (bases 1 to 280143)

## AUTHORS

Wuzy, D. Marie, Metzker, V. Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allon, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, V., Baca, E., Baden, H., Badier, D., Bandaranaike, R., Barber, M., Barnstead, M., Benamoud, F., Bernal, K., Blair, C., Blakenburg, R., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burck, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Ceter, A., Chacko, J., Chavez, B., Chen, G., Chen, R., Chen, Y., Chot, Z., Chu, J., Cleveland, C., Cockrell, P., Cox, C., Foyle, M., Cree, A., D'Souza, D., Davila, M. L., Davis, A., Davy, Carroll, L., De Anda, C., Dederich, P., Delgado, O., Jensen, S., Detaro, C., Ding, Y., Dinh, H., Divya, K., Dreier, B., Dugan-Hochs, S., Dunn, A., Durkin, K., Durval, S., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. V., Gebisi, A., Gant, R., Garcia, A., Garner, T., Garza, M., Gebregiorgis, F., Gess, K., Gill, R., Grady, M., Guerra, M., Guevata, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, J., Hines, S., Hladun, S., Hodgson, A., Hughes, M., Hollins, B., Howells, S., Hui, S., Hui, S., Idrebi, D., Jackson, A., Jackson, L., Jact, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovat, C., Kowals, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, C., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lotens, J., Louised, H., Corado, P. J., Lusk, M., Ma, J., Mathew, B., Mathindane, M., Mahmoud, M., Malloy, K., Mangan, A., Manum, B., Mapou, P., Martin, K., Martin, R., Martinez, F., Maxwell, S., McLeod, M. P., McKelvey, T. Z., Meenen, E., Micosavijevic, A., Mied, G., Naba, E., Montemayor, J., Moore, S., Morgan, V., Morris, K., Morris, S., Yunidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokedemen, C., Okwona, C., Olanrewaju, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, J., Pfannkuch, C., Plopper, F., Poindester, A., Popovic, D., Primus, E., Pritchard, C., Pua, M., Quiroz, C., Racht, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, A., Reilly, M., Reilly, M., Reuter, M., Richards, S., Rivas, C., Rodkey, T., Rojas, A., Rose, X., Rose, A., Ruiz, S. J., Sanders, W., Savary, G., Scherer, S., Scott, G., Scatsman, S., Shen, H., Scott, J., Shvatsky, A., Sisson, I., Sitter, C. F., Smith, D., Speed, A., Sodergren, E., Song, X. Z., Sorrell, R., Sosa, J., Steimle, V., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, K., Thomas, S., Tingey, A., Trejes, A., Usmani, K., Valas, R., Vera, V., Villanar, D., Waldron, L., Walker, R., Wang, Q., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczka, R., Wooden, R., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yen, Y., Yen, Y., Yu, P., Zhang, J., Zhao, X., Zhao, X., Zhao, S., Dunn, D., von Wiersthausen, A., Weiss, E., Smith, D. R., Holt, P. A., Smith, H. O., Weinstock, G., and Gibbs, R. A.

Direct Submission

Unpublished

2 (bases 1 to 280190)

Worley, K. C.

Direct Submission

Submitted (19-FEB-2002)

Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

1 (bases 1 to 280190)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (13-MAY-2003)

Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On May 13, 2003 this sequence version replaced g1:24519078.

The sequence in this assembly is a combination of BAC based reads

and whole genome shotgun sequencing reads assembled using Atlas

(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described

in the feature table below represents a scaffold in the Atlas

assembly (a 'contig-scaffold'). Within each contig-scaffold,

individual sequence contigs are ordered and oriented, and separated

by sized gaps filled with Ns to the estimated size. The sequence

may extend beyond the ends of the clone and there may be sequence

contigs within a contig-scaffold that consist entirely of whole





- \* identifying clones that may be Gene-rich and allows
- \* overlap relationships among clones to be deduced.
- \* However, it should not be assumed that this clone
- \* will be sequenced to completion. In the event that
- \* the record is updated, the accession number will
- \* be preserved.

```

1      5841: contig of 644 bp in length
      663      784: gap of 100 bp
      1490: contig of 726 bp in length
      1593: gap of 100 bp
      1593: contig of 95 bp in length
      2286: gap of 100 bp
      2385: contig of 700 bp in length
      3086      3185: gap of 100 bp
      3185: contig of 691 bp in length
      3278: gap of 100 bp
      3278: contig of 726 bp in length
      4684: gap of 100 bp
      4754: gap of 100 bp
      5492: contig of 498 bp in length
      5582: gap of 100 bp
      5583: contig of 705 bp in length
      6287: gap of 100 bp
      7078: contig of 691 bp in length
      7178: gap of 100 bp
      7881: contig of 703 bp in length
      7981: gap of 100 bp
      8653      8752: gap of 100 bp
      9446: contig of 694 bp in length
      9546: gap of 100 bp
      10221: contig of 475 bp in length
      10221: gap of 100 bp
      10323: contig of 702 bp in length
      11024      11123: gap of 100 bp
      11123: contig of 695 bp in length
      11321      11320: gap of 100 bp
      11321: contig of 692 bp in length
      11921      12712: gap of 100 bp
      12712: contig of 682 bp in length
      12673      13194: contig of 695 bp in length
      13195      13493: gap of 100 bp
      13493: contig of 696 bp in length
      14193      14293: gap of 100 bp
      14293: contig of 705 bp in length
      14396: gap of 100 bp
      14397: gap of 100 bp
      15089: contig of 693 bp in length
      15792      15889: gap of 100 bp
      15890      16500: contig of 611 bp in length
      16501      16601: gap of 100 bp
      16601: contig of 695 bp in length
      17296      17396: gap of 100 bp
      17396: contig of 692 bp in length
      18088      18187: gap of 100 bp
      18188      18887: contig of 703 bp in length
      18888      19681: gap of 100 bp
      19681: contig of 694 bp in length
      19682      19781: gap of 100 bp
      19782      20470: contig of 689 bp in length
      20471      20570: gap of 100 bp
      20571      21135: contig of 565 bp in length
      21136      21236: gap of 100 bp
      21236      21920: contig of 695 bp in length
      21921      22020: gap of 100 bp
      22021      22720: contig of 704 bp in length
      22721      22820: gap of 100 bp
      22821      23526: contig of 702 bp in length
      23527      23626: gap of 100 bp
      23627      24309: contig of 683 bp in length
      24310      24409: gap of 100 bp
      24410      25107: contig of 699 bp in length
      25208      25207: gap of 100 bp
      25209      25874: contig of 667 bp in length
      25875      25974: gap of 100 bp
      25975      26682: contig of 708 bp in length

```

```

26782      27463: gap of 100 bp
      27463: contig of 681 bp in length
      27561: gap of 100 bp
      28251: contig of 690 bp in length
      28353: gap of 100 bp
      29046: contig of 693 bp in length
      29146: gap of 100 bp
      29817      29817: contig of 691 bp in length
      29818      29917: gap of 100 bp
      29918      30625: contig of 688 bp in length
      30626      31476: gap of 100 bp
      31476: contig of 681 bp in length
      31574      32186: gap of 100 bp
      32186: contig of 680 bp in length
      32286      32984: gap of 100 bp
      32984: contig of 698 bp in length
      33084      33775: gap of 100 bp
      33775: contig of 691 bp in length
      33875      34588: gap of 100 bp
      34588: contig of 693 bp in length
      34688      35355: gap of 100 bp
      35355: contig of 687 bp in length
      35455      36164: gap of 100 bp
      36164: contig of 709 bp in length
      36264      36963: gap of 100 bp
      36963: contig of 699 bp in length
      37063      37759: gap of 100 bp
      37759: contig of 696 bp in length
      37859      38548: gap of 100 bp
      38548: contig of 689 bp in length
      38648      39351: gap of 100 bp
      39351: contig of 702 bp in length
      39451      40134: gap of 100 bp
      40134: contig of 684 bp in length
      40234      40948: gap of 100 bp
      40948: contig of 704 bp in length
      41048      41738: gap of 100 bp
      41738: contig of 700 bp in length
      41838      42543: gap of 100 bp
      42543: contig of 705 bp in length
      42643      43332: gap of 100 bp
      43332: contig of 689 bp in length
      43432      44117: gap of 100 bp
      44117: contig of 685 bp in length
      44217      44908: gap of 100 bp
      44908: contig of 691 bp in length
      45008      45715: gap of 100 bp
      45715: contig of 707 bp in length
      45815      46507: gap of 100 bp
      46507: contig of 692 bp in length
      46607      47294: gap of 100 bp
      47294: contig of 687 bp in length
      47394      48096: gap of 100 bp
      48096: contig of 702 bp in length
      48196      48901: gap of 100 bp
      48901: contig of 705 bp in length
      49001      49702: gap of 100 bp
      49702: contig of 699 bp in length
      49802      49900: gap of 100 bp
      49900: contig of 684 bp in length
      50084      50485: gap of 100 bp
      50485: contig of 691 bp in length
      50585      51275: gap of 100 bp
      51275: contig of 700 bp in length
      51376      52175: gap of 100 bp
      52175: contig of 676 bp in length
      52276      52851: gap of 100 bp
      52851: contig of 698 bp in length
      52952      53749: gap of 100 bp
      53749: contig of 695 bp in length
      53850

```

Query Match: 1.7% Score 22; DB 2; Length 57698;  
 Best Local Similarity: 100.0%; Pred. No. 2.2;

```

Matches 22: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1142 AGGATCCTTAGGTCGCTGA 1163
DB 7640 AGGATCCTTAGGTCGCTGA 7661
|||||
|||||

RESULT 34
AC090374 61859 bp DNA Linear HTG 22 MAR 2001
LOCUS Homo sapiens chromosome 18 clone RP11-776M20 map 18, LOW PASS
DEFINITION AC090374
ACCESSION AC090374
VERSION G1:13411051
KEYWORDS HTG; HTGS PHASE0
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumetazoa;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1. Buresi L TO 61859;
AUTHORS Birren,B., Linton,L., Nusbaum,C and Lander,E.
TITLE Homo sapiens chromosome 18, clone RP11-776M20
JOURNAL Unpublished
2. Buresi L TO 61859;
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Batina,N., Bustier,V., Boguslavsky,L., Boukhgalter,B., Brown,A.,
Camarata,J., Campopiano,A., Czeipel,Y., Collangelo,M., Collins,S.,
Collins,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Fero,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,K.,
Hagos,B., Harford,A., Horton,C., Hueme,W., Iliev,I., Johnson,R.,
Jones,C., Karafas,A., LaRocque,K., Lamazates,R., Lander,T.,
Lehoczky,S., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPherson,R., Meldrum,S., Meneses,L., Mohovatz,T., Miengo,V.,
Murphy,T., Naylor,C., Nguyen,C., Nodou,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,C., Peterson,K.,
Phukhuth,P., Pierre,K., Pollara,V., Raymond,C., Pettar,R.,
Ritback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schaefer,S., Schuback,R., Seaman,S., Sevely,P.,
Sougnuez,C., Spencer,B., Stange,Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Tadmor,J., Tesfaye,S., Theodore,J.,
Tsvetkov,M., Travis,N., Triggiano,J., Vassiliev,R., Veldin,R., Volin,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,S., Zehnoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
DIRECT SUBMISSION
TITLE Direct Submission
JOURNAL Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Mar 22, 2001 this sequence version replaced g1:12958010.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997).
http://ftp.genome.washington.edu/RM/RepeatMasker.htm
..... Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
..... Project Information
Center project name: L12667
Center clone name: 776_M_20
.....
NOTE: This record contains 78 individual:
• sequencing reads that have not been assembled into
contigs. Runs of N are used to separate the reads
• and the order in which they appear is completely
arbitrary. Low-pass sequence sampling is useful for
identifying clones that may be gene-rich and allows
overlap relationships among clones to be deduced.
• However, it should not be assumed that this clone
will be sequenced to completion. In the event that
the record is updated, the accession number will
be preserved.
1 680: contig of 680 bp in length
2 781 1383: contig of 703 bp in length
3 1584 2313: contig of 730 bp in length
4 2414 3115: contig of 702 bp in length
5 3116 3497: contig of 682 bp in length
6 3498 4356: contig of 689 bp in length
7 4357 5480: contig of 694 bp in length
8 5481 6170: contig of 692 bp in length
9 6171 7661: contig of 695 bp in length
10 7662 8067: contig of 686 bp in length
11 8068 8567: contig of 684 bp in length
12 8568 10346: contig of 679 bp in length
13 10347 11451: contig of 699 bp in length
14 11452 11950: contig of 705 bp in length
15 11951 12623: contig of 673 bp in length
16 12624 13481: contig of 678 bp in length
17 13482 14581: contig of 686 bp in length
18 14582 14771: contig of 684 bp in length
19 14772 15742: contig of 709 bp in length
20 15743 16376: contig of 699 bp in length
21 16377 17423: contig of 686 bp in length
22 17424 18521: contig of 683 bp in length
23 18522 18941: contig of 689 bp in length
24 18942 19341: contig of 737 bp in length
25 19342 20433: contig of 605 bp in length
26 20434 21326: contig of 683 bp in length
27 21327 22115: contig of 699 bp in length
28 22116 23068: contig of 693 bp in length
29 23069 23755: contig of 681 bp in length
30 23756 24377: contig of 688 bp in length
31 24378 25276: contig of 699 bp in length
32 25277 26357: contig of 681 bp in length
33 26358 26556: contig of 693 bp in length
34 26557 26951: contig of 700 bp in length
35 26952 27750: contig of 706 bp in length
36 27751 28556: contig of 685 bp in length
37 28557 29441: contig of 703 bp in length
38 29442

```

```

• 20142 30528: contig of 497 bp in length
• 30029 30124: gap of 170 bp
• 30124 30144: contig of 496 bp in length
• 30144 30159: gap of 100 bp
• 30159 31598: contig of 64 bp in length
• 31598 31698: gap of 100 bp
• 31698 32497: contig of 479 bp in length
• 32497 33190: gap of 100 bp
• 33190 33290: contig of 403 bp in length
• 33290 33296: gap of 100 bp
• 33296 33996: contig of 736 bp in length
• 33996 34096: gap of 100 bp
• 34096 34771: contig of 675 bp in length
• 34771 34871: gap of 100 bp
• 34871 35593: contig of 313 bp in length
• 35593 35689: gap of 100 bp
• 35689 36393: contig of 700 bp in length
• 36393 36499: gap of 100 bp
• 36499 37192: contig of 723 bp in length
• 37192 37292: gap of 100 bp
• 37292 37957: contig of 665 bp in length
• 37957 38257: gap of 100 bp
• 38257 38554: contig of 637 bp in length
• 38554 38554: gap of 100 bp
• 38554 39331: contig of 677 bp in length
• 39331 39331: gap of 100 bp
• 39331 40333: contig of 702 bp in length
• 40333 40333: gap of 100 bp
• 40333 41145: contig of 712 bp in length
• 41145 41245: gap of 100 bp
• 41245 41263: contig of 441 bp in length
• 41263 42263: gap of 100 bp
• 42263 42425: contig of 639 bp in length
• 42425 42425: gap of 100 bp
• 42425 43511: contig of 686 bp in length
• 43511 43611: gap of 100 bp
• 43611 44233: contig of 682 bp in length
• 44233 44333: gap of 100 bp
• 44333 45283: contig of 690 bp in length
• 45283 45283: gap of 100 bp
• 45283 45886: contig of 725 bp in length
• 45886 45986: gap of 100 bp
• 45986 46672: contig of 686 bp in length
• 46672 46772: gap of 100 bp
• 46772 47489: contig of 617 bp in length
• 47489 47489: gap of 100 bp
• 47489 48285: contig of 614 bp in length
• 48285 48285: gap of 100 bp
• 48285 48294: contig of 614 bp in length
• 48294 48294: gap of 100 bp
• 48294 49093: contig of 615 bp in length
• 49093 50059: gap of 100 bp
• 50059 50714: contig of 703 bp in length
• 50714 50814: gap of 100 bp
• 50814 51499: contig of 645 bp in length
• 51499 51599: gap of 100 bp
• 51599 52290: contig of 691 bp in length
• 52290 52903: gap of 100 bp
• 52903 53077: contig of 687 bp in length
• 53077 53177: gap of 100 bp
• 53177 53660: contig of 683 bp in length
• 53660 53963: gap of 100 bp
• 53963 54675: contig of 715 bp in length

```

Query Match Best Local Similarity 100.0%, Score 22, DB 2, Length 61859, Matches 22, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY 1142 AGGATCCCTTAGGTCATGCTGA 1163

DB 1408 AGGATCCCTTAGGTCATGCTGA 1429

## RESULT IS

AC090374

LOCUS

DEFINITION

AC090374

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AC090374 Homo sapiens chromosome 18 clone RP11-776M20 map 18, LCM-PASS  
SEQUENCE SAMPLING.

AC090374 AC090374.2 GI:1347105;  
HTG: HTGS PHASE3  
Homo sapiens (human)

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 61859)  
Britten, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,  
Barns, N., Bastien, V., Boeke, J., Brown, A., Brown, A., Brown, A.,  
Carrara, J., Campotondo, A., Choquet, Y., Colangelo, M., Collins, S.,  
Collins, A., Cooke, P., DeArnell, K., Dewar, K., Diaz, J.,  
Dodge, S., Fero, S., Ferrante, P., FitzHugh, K., Gage, J., Galagan, J.,  
Gardner, S., Ginde, S., Gysler, M., Grubbs, L., Grubbs, L., Grubbs, L.,  
Hagood, J., Harford, A., Horton, L., Hulme, W., Iliev, I., Johnson, S.,  
Jones, C., Karas, A., Laskowski, K., Lazarides, P., Landers, J.,  
Lefcowsky, J., Levine, R., Liu, S., MacLean, C., Macdonald, P.,  
McGee, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,  
McKernan, R., McQuinn, J., Meneus, L., Minova, T., Miska, V.,  
Murphy, T., Nay, C., Nguyen, C., Norbu, C., Norman, C.,  
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
Phunkang, P., Pierre, N., Pollara, V., Raymond, C., Reilly, M.,  
Reilly, M., Piley, R., Piley, C., Rogov, P., Roman, S., Rosetti, M.,  
Roy, A., Saito, R., Schauer, S., Schupbach, R., Seaman, S., Seavey, B.,  
Sennarath, C., Spencer, B., Stange, Thomas, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talaras, J., Tesfaye, S., Theodore, J.,  
Travers, M., Travis, N., Triggiani, J., Vassiliev, H., Viel, R., VO, A.,  
Wilson, M., Wu, X., Wyman, S., Ye, W., Young, G., Zainuddin, J.,  
Zerkow, L., Zimmer, A. and Zody, M.

2 (bases 1 to 61859)  
Britten, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,  
Barns, N., Bastien, V., Boeke, J., Brown, A., Brown, A., Brown, A.,  
Carrara, J., Campotondo, A., Choquet, Y., Colangelo, M., Collins, S.,  
Collins, A., Cooke, P., DeArnell, K., Dewar, K., Diaz, J.,  
Dodge, S., Fero, S., Ferrante, P., FitzHugh, K., Gage, J., Galagan, J.,  
Gardner, S., Ginde, S., Gysler, M., Grubbs, L., Grubbs, L., Grubbs, L.,  
Hagood, J., Harford, A., Horton, L., Hulme, W., Iliev, I., Johnson, S.,  
Jones, C., Karas, A., Laskowski, K., Lazarides, P., Landers, J.,  
Lefcowsky, J., Levine, R., Liu, S., MacLean, C., Macdonald, P.,  
McGee, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,  
McKernan, R., McQuinn, J., Meneus, L., Minova, T., Miska, V.,  
Murphy, T., Nay, C., Nguyen, C., Norbu, C., Norman, C.,  
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
Phunkang, P., Pierre, N., Pollara, V., Raymond, C., Reilly, M.,  
Reilly, M., Piley, R., Piley, C., Rogov, P., Roman, S., Rosetti, M.,  
Roy, A., Saito, R., Schauer, S., Schupbach, R., Seaman, S., Seavey, B.,  
Sennarath, C., Spencer, B., Stange, Thomas, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talaras, J., Tesfaye, S., Theodore, J.,  
Travers, M., Travis, N., Triggiani, J., Vassiliev, H., Viel, R., VO, A.,  
Wilson, M., Wu, X., Wyman, S., Ye, W., Young, G., Zainuddin, J.,  
Zerkow, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 22, 2001 this sequence version replaced gi1347105.1.  
All repeats were identified using RepeatMasker:  
Smith, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/PW/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIPR  
Web site: <http://www.sage.wi.mit.edu>

Contact: sequence\_submission@genome.wi.mit.edu  
Project Information  
Center project name: L12657  
Center clone name: 776\_M\_20

NOTE: This record contains 78 individual  
sequencing reads that have not been assembled into  
contigs. Runs of N are used to separate the reads  
and the order in which they appear is completely  
arbitrary. Low pass sequence sampling is useful for  
identifying clones that may be gene-rich and allows  
overlap relationships among clones to be deduced.  
However, it should not be assumed that this clone  
will be sequenced to completion. In the event that  
the record is updated, the accession number will  
be preserved.

1 680: contig of 680 bp in length  
681 780: gap of 100 bp  
781 1483: contig of 703 bp in length  
1484 1584: gap of 100 bp  
1584 2314: contig of 730 bp in length  
2314 2414: gap of 100 bp  
2414 3115: contig of 702 bp in length  
3115 3215: gap of 100 bp





```

35265 35977: contig of 713 bp in length
35978 36077: gap of 100 bp
36079 36966: contig of 719 bp in length
36970 36996: gap of 100 bp
36997 37594: contig of 698 bp in length
37595 37694: gap of 100 bp
37695 38375: contig of 682 bp in length
38376 38475: gap of 100 bp
38476 39177: contig of 702 bp in length
39178 39277: gap of 100 bp
39278 39985: contig of 708 bp in length
39986 40086: gap of 100 bp
40087 40798: contig of 713 bp in length
40799 40898: gap of 100 bp
40899 41583: contig of 684 bp in length
41584 41683: gap of 100 bp
41684 42367: contig of 685 bp in length
42368 42467: gap of 100 bp
42468 43158: contig of 689 bp in length
43159 43258: gap of 100 bp
43259 43963: contig of 707 bp in length
43964 44063: gap of 100 bp
44064 44768: contig of 705 bp in length
44769 44868: gap of 100 bp
44869 45541: contig of 673 bp in length
45542 45641: gap of 100 bp
45642 46311: contig of 670 bp in length
46312 46411: gap of 100 bp
46412 47106: contig of 695 bp in length
47107 47206: gap of 100 bp
47207 47903: contig of 699 bp in length
47904 48003: gap of 100 bp
48004 48691: contig of 686 bp in length
48692 48791: gap of 100 bp
48792 49492: contig of 701 bp in length
49493 49592: gap of 100 bp
49593 50293: contig of 701 bp in length
50294 50393: gap of 100 bp
50394 51112: contig of 719 bp in length
51113 51212: gap of 100 bp
51213 51928: contig of 716 bp in length
51929 52028: gap of 100 bp
52029 52741: contig of 713 bp in length
52742 52841: gap of 100 bp
52842 53533: contig of 692 bp in length
53534 53633: gap of 100 bp
53634 54332: contig of 699 bp in length
54333 54432: gap of 100 bp
54433 55138: contig of 706 bp in length

```

```

Query Match: 1.71; Score 22; EB 2; Length 64693;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1142 AGGATCGCTAGCTTGATGCTGA 1163
|||||
DB 5844 AGGATCGCTAGCTTGATGCTGA 5865

```

```

RESULT J7
AC105141/c AC105141 64693 bp DNA linear HTG 23 DEC-2001
DEFINITION Homo sapiens chromosome 18 clone RP11-34312 map 16, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC105143
VERSION AC105143.1 GI:17977639
KEYWORDS HTG; HTGS PHASE0.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 64693)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

```

# TITLE JOURNAL REFERENCE AUTHORS

Homo sapiens chromosome 18, clone RP11-34312

Unpublished  
2 (bases 1 to 64693)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Boguski,M.S., Bouckgeer,A.,  
Brown,A., Cantarel,D., Campolongo,A., Chang,J., Chazotte,B.,  
Choquet,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,  
Conceicao,P., DeAranda,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S.,  
Ferreira,P., FitzHugh,W., Gale,E., Galagan,J., Gardy,J.,  
Glenn,S., Goto,S., Guyette,M., Guzik,L., Grand-Pierre,N.,  
Hague,R., Heaford,A., Holton,J., Hume,W., Iliev,I., Johnson,R.,  
Jones,C., Khatami,A., Karasik,A., Kelis,C., Larocque,K.,  
Lamarche,P., Landers,T., Lenczky,J., Levine,R., Liu,G.,  
Maclean,C., Mardona,J.P., Maizels,N., Marquis,N., Matthews,C.,  
McCarthy,M., McKean,P., McKinnon,K., McPheters,R., Moldavi,C.,  
Meredith,J., Mihova,T., Miengo,V., Murphy,T., Nayler,J., Nguyen,C.,  
Notha,C., Norman,H., O'Connor,T., O'Donnell,P., O'Neill,D.,  
Olivero,A., Peterson,K., Pflueger,B., Pierre,N., Pollack,V.,  
Raymond,R., Rife,P., Rhee,M., Riley,K., Rose,C., Rogov,P.,  
Roman,R., Rong,Y., Roy,A., Saito,K., Schaefer,T., Schuppach,R.,  
Seaman,R., Sherry,P., Spake,R., Stange,Thomann,N., Stancovski,N.,  
Stevens,R., Subramanian,A., Tamas,J., Tesfaye,S., Theodore,C.,  
Tobin,K., Travers,M., Travis,N., Tzeng,L., Vassiliev,H.,  
Victor,R., Voigt,W., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,S.,  
Zandbergen,C., Zembek,D., Zimmer,A. and Zody,M.

Direct Submission

# TITLE JOURNAL

# COMMENT

Submitted 123-DEC-2001; Whitehead Institute/MIT Center for Genome  
Repeats, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Star/A.F.A. A Green, 11/09/1998  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIR2

Web site: http://www.seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

Center project name: 223858

Center clone name: 34 B 12

NOTE: This record contains 81 individual  
sequencing reads that have not been assembled into  
contigs. Runs of N are used to separate the reads  
and the order in which they appear is completely  
arbitrary. Now pars sequence sampling is useful for  
identifying clones that may be generic and allows  
overlap relationships among clones to be deduced.  
However, it should not be assumed that this clone  
will be sequenced to completion. In the event that  
the record is updated, the accession number will  
be preserved.

706: contig of 756 bp in length  
707 806: gap of 100 bp  
807 1512: contig of 755 bp in length  
1513 1612: gap of 100 bp  
1613 2425: contig of 713 bp in length  
2426 2425: gap of 100 bp  
2426 3108: contig of 693 bp in length  
3109 3708: gap of 100 bp  
3709 3928: contig of 720 bp in length  
3929 4328: gap of 100 bp  
4329 4743: contig of 715 bp in length  
4744 4843: gap of 100 bp  
4844 5548: contig of 705 bp in length  
5549 5648: gap of 100 bp  
5649 6351: contig of 703 bp in length  
6352 6451: gap of 100 bp  
6452 7148: contig of 697 bp in length  
7149 7248: gap of 100 bp  
7249 7949: contig of 701 bp in length  
7950 8049: gap of 100 bp  
8050 8743: contig of 694 bp in length  
8744 8843: gap of 100 bp

8844 9311: contig of 688 bp in length  
 9532 9611: gap of 100 bp  
 9532 10126: contig of 695 bp in length  
 13327 10426: gap of 100 bp  
 13427 11124: contig of 698 bp in length  
 13427 11224: gap of 100 bp  
 13427 11336: contig of 712 bp in length  
 13427 12336: gap of 100 bp  
 13427 12446: contig of 710 bp in length  
 13427 13566: contig of 710 bp in length  
 13427 13656: gap of 100 bp  
 13427 14176: contig of 719 bp in length  
 13427 14476: gap of 100 bp  
 13427 15162: contig of 707 bp in length  
 13427 15262: gap of 100 bp  
 13427 15366: contig of 703 bp in length  
 13427 16065: gap of 100 bp  
 13427 16167: contig of 702 bp in length  
 13427 16667: gap of 100 bp  
 13427 17667: contig of 703 bp in length  
 13427 17667: gap of 100 bp  
 13427 18352: contig of 685 bp in length  
 13427 18452: gap of 100 bp  
 13427 19152: contig of 710 bp in length  
 13427 19355: contig of 713 bp in length  
 13427 20295: gap of 100 bp  
 13427 22000: gap of 100 bp  
 13427 22599: contig of 699 bp in length  
 13427 22800: gap of 100 bp  
 13427 22922: gap of 100 bp  
 13427 24196: contig of 704 bp in length  
 13427 24286: gap of 100 bp  
 13427 24395: contig of 699 bp in length  
 13427 24793: contig of 698 bp in length  
 13427 24893: gap of 100 bp  
 13427 25582: contig of 690 bp in length  
 13427 25682: gap of 100 bp  
 13427 26193: contig of 716 bp in length  
 13427 26494: gap of 100 bp  
 13427 27221: contig of 721 bp in length  
 13427 27321: gap of 100 bp  
 13427 28116: contig of 704 bp in length  
 13427 28116: gap of 100 bp  
 13427 28213: contig of 698 bp in length  
 13427 28213: gap of 100 bp  
 13427 29702: gap of 100 bp  
 13427 30397: contig of 695 bp in length  
 13427 30497: gap of 100 bp  
 13427 31205: contig of 708 bp in length  
 13427 31305: gap of 100 bp  
 13427 31358: contig of 693 bp in length  
 13427 32058: gap of 100 bp  
 13427 32749: contig of 691 bp in length  
 13427 32849: gap of 100 bp  
 13427 33551: contig of 702 bp in length  
 13427 33551: gap of 100 bp  
 13427 34356: contig of 705 bp in length  
 13427 34456: gap of 100 bp  
 13427 35164: contig of 728 bp in length  
 13427 35264: gap of 100 bp  
 13427 35977: contig of 713 bp in length  
 13427 36077: gap of 100 bp  
 13427 36796: contig of 719 bp in length  
 13427 36896: gap of 100 bp  
 13427 37594: contig of 699 bp in length  
 13427 37694: gap of 100 bp  
 13427 38175: contig of 681 bp in length

38475: gap of 100 bp  
 39177: contig of 702 bp in length  
 39177: gap of 100 bp  
 39885: contig of 728 bp in length  
 40085: gap of 100 bp  
 40798: contig of 713 bp in length  
 40799: gap of 100 bp  
 41582: contig of 684 bp in length  
 41582: gap of 100 bp  
 42167: contig of 685 bp in length  
 42467: gap of 100 bp  
 43156: contig of 689 bp in length  
 43157: gap of 100 bp  
 43963: contig of 707 bp in length  
 44064: gap of 100 bp  
 44768: contig of 705 bp in length  
 44769: gap of 100 bp  
 45441: contig of 673 bp in length  
 45442: gap of 100 bp  
 45642: contig of 670 bp in length  
 46111: gap of 100 bp  
 47106: contig of 695 bp in length  
 47206: gap of 100 bp  
 47933: contig of 699 bp in length  
 48063: gap of 100 bp  
 48691: contig of 686 bp in length  
 48791: gap of 100 bp  
 49492: contig of 701 bp in length  
 49492: gap of 100 bp  
 50293: contig of 701 bp in length  
 50393: gap of 100 bp  
 51113: contig of 719 bp in length  
 51212: gap of 100 bp  
 51928: contig of 716 bp in length  
 52028: gap of 100 bp  
 52741: contig of 713 bp in length  
 52941: gap of 100 bp  
 52941: contig of 692 bp in length  
 53533: gap of 100 bp  
 54334: contig of 699 bp in length  
 54334: gap of 100 bp  
 55138: contig of 716 bp in length

Query Match  
 Best Local Similarity 100.0%, Pred. No. 2.2,  
 Matches 22, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

CY 1147 AGGATCTTACCTTATCTCTCA 1163  
 CT 1146 AGGATCTTACCTTATCTCTCA 19185

RESULT 3A  
 AC079625.0  
 WPCOMMENT

Sequence split into 4 fragments LOCUS AC079625 Accession AC079625  
 Fragment Name Begin End  
 AC079625.0 1 110000  
 AC079625.1 100001 210000  
 AC079625.2 200001 310000  
 AC079625.3 300001 352210

LOCUS AC079625 352210 bp DNA linear HTG 25-AUG-2002  
 DEFINITION Homo sapiens chromosome 12 clone RP11-657D11, WORKING DRAFT  
 SEQUENCE, 7 unordered pieces.

ACCESSION AC079625  
 VERSION AC079625.18 GI:22138234  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 352210)  
 AUTHORS Muzny,D.M., Adams,C., Adio-Gduola,B., Ali-osman,F.R., Allen,C.,









```

* 124280 contig of 5039 bp in length
* 129319 gap of 100 bp
* 133871 contig of 4459 bp in length
* 134338 gap of 100 bp
* 137358 contig of 3421 bp in length
* 137359 gap of 100 bp
* 137459 contig of 1478 bp in length
* 140917 contig of 100 bp
* 141037 gap of 100 bp
* 142151 contig of 1115 bp in length
* 142152 gap of 100 bp
* 142352 contig of 1113 bp in length

```

## FEATURES

```

Location/Qualifiers
1..143364
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chrchromosomes="18"
/map="18q12"
/clone="RP11-776B3"
1..46191
/note="assembly_fragment"
46482..53705
/note="assembly_fragment"
53806..69654
/note="assembly_fragment"
69755..89552
/note="assembly_fragment"
89653..103532
/note="assembly_fragment"
103633..114843
/note="assembly_fragment"
114944..124179
/note="assembly_fragment"
124280..124318
/note="assembly_fragment"
129419..133837
/note="assembly_fragment"
133938..137356
/note="assembly_fragment"
137459..140936
/note="assembly_fragment"
141037..142151
/note="assembly_fragment"
142252..143364
/note="assembly_fragment"
46280 a 29927 c 27209 g 46246 t 1200 others
ORIGIN

```

```

Query Match 1 354 Store 100 100 100 100 100 100 100 100 100 100
Best local similarity 100.00, Prod. 100.00, 100.00, 100.00, 100.00, 100.00, 100.00, 100.00, 100.00, 100.00
Matches 22: Conservative 0; Mismatches 0; Indels 0; Gaps 0

```

```

QY 1142 AGGATCTTGGTTCATGCTGA 1143
|||||
DB 14475 AGGATCTTGGTTCATGCTGA 44454
|||||

```

```

RESULT 42
AC090373 143456 bp CNA linear PRI 16-MAR-2003
DEFINITION Homo sapiens chromosome 18, clone RP11-776B3, complete sequence.
ACCESSION AC090373
VERSION 1.1
KEYWORDS HGT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 143456)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 18, clone RP11-776B3
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 143456)

```

## AUTHORS

Birren,B., Linton,J., Nusbaum,C., Lander,E., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavsky,I., Bouckgeater,B., Brown,A., Cantata,A., Campopiano,A., Choquet,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArliano,K., Dewar,K., Diaz,C.S., Dodge,S., Fato,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Girde,S., Goyette,M., Graham,L., Grand-Pierre,K., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRoque,K., Larazares,R., Landers,T., Lenczyk,J., Levine,R., Liu,G., Maclean,C., MacDonald,P., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKnight,K., McPherson,R., Melidrim,J., Menes,L., Mihova,T., Mironov,V., Murphy,T., Nay,O., Nguyen,C., North,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Piorie,N., Pollara,V., Raymond,C., Retta,B., Rieback,M., Riley,P., Rise,C., Rogov,P., Roman,C., Rosetti,M., Roy,A., Santos,R., Schauer,B., Schupack,R., Seaman,S., Severy,P., Senguez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Travers,M., Travis,N., Triglilio,C., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W., Young,G., Zainoun,J., Zembek,B., Zimmer,A. and Zody,M.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

Submitted (17-FEB-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 143456)  
 Birren,B., Nusbaum,C., Lander,E., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,I., Bouckgeater,B., Camarata,J., Chang,J., Chazaro,B., Choquet,Y., Collymore,A., Cooke,A., Cooke,P., DeArliano,K., Dewar,K., Diaz,C.S., Dodge,S., Fato,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Gird,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Karat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., MacDonald,P., Naylor,J., Nguyen,C., McCarthy,M., Melidrim,J., Menes,L., Mihova,T., Mironov,V., Murphy,T., Nay,O., Nguyen,C., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Piorie,N., Pollara,V., Raymond,C., Retta,B., Rise,C., Rogov,P., Roman,C., Roy,A., Schauer,B., Schupack,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Travers,M., Travis,N., Triglilio,C., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,B., Zimmer,A. and Zody,M.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

Submitted (24-DEC-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 4 (bases 1 to 143456)  
 Birren,B., Nusbaum,C., Lander,E., Abouelheil,A., Allen,N., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavsky,I., Bouckgeater,B., Camarata,J., Chang,C., Choquet,Y., Collymore,A., Cooke,A., Cooke,P., Corum,B., DeArliano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorts,L., Erickson,J., Fato,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Karat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Liu,A., Mabbitt,R., Maclean,C., MacDonald,P., Major,S., Manning,C., Matthews,C., McCarthy,M., Melidrim,J., Menes,L., Mihova,T., Mironov,V., Murphy,T., Nay,O., Nguyen,C., Nicol,R., North,C., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Piorie,N., Rieback,M., Ramasamy,U., Raymond,C., Retta,B., Rise,C., Rogov,P., Roman,C., Santos,R., Schupack,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataratan,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,B., Zimmer,A. and Zody,M.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

Submitted (06-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Mar 6, 2003 this sequence version replaced gi:27369463.  
 All repeats were identified using RepeatMasker.  
 Smit, A.F.A. & Green, P. (1996-1997)





















```

1 46151 46250: gap of 100 bp
1 47227: contig of 147 bp in length
1 47827: gap of 100 bp
1 48840: contig of 102 bp in length
1 49442: gap of 100 bp
1 48941 50133: contig of 1415 bp in length
1 50136 50459: gap of 100 bp
1 50459 51763: contig of 1354 bp in length
1 51764 51843: gap of 100 bp
1 51843 53443: contig of 1480 bp in length
1 53443 54043: gap of 100 bp
1 54043 55043: contig of 1569 bp in length
1 55043 55143: gap of 100 bp
1 55143 56534: contig of 1354 bp in length
1 56534 56634: gap of 100 bp
1 56634 57911: contig of 1277 bp in length
1 57911 58011: gap of 100 bp
1 58011 58911: contig of 900 bp in length
1 58911 59211: gap of 100 bp
1 59211 60123: contig of 1112 bp in length
1 60123 60223: gap of 100 bp
1 60223 61343: contig of 1120 bp in length
1 61343 61443: gap of 100 bp
1 61443 63343: contig of 1903 bp in length
1 63343 64443: gap of 100 bp
1 64443 64543: contig of 800 bp in length
1 64543 64643: gap of 100 bp
1 64643 66553: contig of 2105 bp in length
1 66553 67443: gap of 100 bp
1 67443 67543: contig of 1194 bp in length
1 67543 67643: gap of 100 bp
1 67643 68558: contig of 1710 bp in length
1 68558 69558: gap of 100 bp
1 69558 70839: contig of 1181 bp in length
1 70839 70939: gap of 100 bp
1 70939 71943: contig of 973 bp in length
1 71943 72043: gap of 100 bp
1 72043 73843: contig of 1870 bp in length
1 73843 73943: gap of 100 bp
1 73943 75943: contig of 1924 bp in length
1 75943 76043: gap of 100 bp

Query Match
Best Local Similarity 100.0% (Fixed No. 24)
Matches 24 Mismatches 0 Indels 0 Gaps 0

CY 541 GAGTAAATGTCAGATGATTTG
|||||
DR 150688 GAGCAATGTCAGATGATTTG

RESULT 12
ACCESSION AC115252 227695 bp ENA 13-NOV-2002
DEFINITION Rattus norvegicus clone CH230-24992, WORKING DRAFT SEQUENCE, 4
uncloned pieces.
ACCESSION AC115252
VERSION AC115252.4 G124941876
KEYWORDS HTG: HTGS PHASE1: HTGS DRAFT: HTGS FULLTOP
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 227695)
AUTHORS Vuzny D, Marie J, Mezker M, Lee J, Abramson S, Aims G, Ader J, Allen C, Allen H, Alsbrooks S, Amin A, Aquilano D, Anyalebechi V, Aoyagi A, Ayodeji K, Bacallao R, Baldwin D, Bandaranaike B, Batbet M, Barrstead M, Benhabib P, Biswa C, Blair C, Blankenburg K, Blythe P, Brown M, Bryant N, Buhay C, Burch P, Butelli K, Calgeron E, Cardenas V, Carter K, Cavazos J, Ceasar H, Center A, Chacko J, Chavez D, Chen G, Chen R, Chen Y, Chen Z, Chu G,

```

```

Cleveland C, Cockrell R, Cox C, Coyte M, Cree A, D'Souza L, Davila M, Davis C, Davy Carroll L, De Anda C, Federich D, Delgado C, Denison S, DeRamo C, Ding Y, Dirlikov K, Divya K, Draper H, Dugan-Rocha S, Dunn A, Durbin K, Duva B, Faves K, Fgan A, Escotto M, Eugene C, Evans C, Falls T, Fan G, Fernandez S, Finley V, Flagg N, Forbes S, Foster M, Foster P, Fraser C, Gail R, Garcia A, Garner T, Garza M, Gebregregios E, Gess K, Gill R, Grady M, Guerra W, Guevara M, Gunaratne P, Hauland K, Hail C, Hamilton C, Hamilton K, Harvey Y, Havlak F, Hawes A, Henderson N, Hernandez J, Hernandez R, Hines S, Hladik S, Hodgson A, Hughes M, Hollins R, Howells S, Huiy K, Hume S, Idebird D, Jackson A, Jackson B, Jacobson J, Jiang H, Johnson B, Johnson R, Johnson A, Karpathy S, Kelly S, Kelly S, Khan Z, King L, Kovar C, Kowis C, Kraft C, Lebow H, Levan J, Lewis L, Li Z, Li J, Liu J, Liu W, Liu Y, London P, Longacre S, Lopez J, Lorenshewa L, Louissegh H, Lozado R, Lu X, Ma J, Maheshwari M, Mahindaratne M, Mahmoud M, Malloy K, Mangum A, Mangum B, Mapua P, Martin K, Martin R, Martinez E, Mathew S, McLeod M, McKell T, Meene E, Milosavljevic A, Vior G, Winta E, Montemayor J, Moore S, Morgan M, Morris K, Morris S, Munidasa M, Murphy M, Nait J, Nankervis C, Neal W, Newton N, Nguyen K, Norris S, Nwakwelench O, Okwuonu G, Olarnpusagoon A, Pal S, Parks K, Pastorkak S, Paul H, Perez A, Perez L, Prankoch C, Pieper F, Poiroux A, Popovic D, Pramus E, Pu L, Puazo X, Quirroz R, Rachlin E, Reeves K, Regier X, Reigler R, Reilly B, Reilly M, Ren Y, Reuter X, Richards S, Riggs P, Rives C, Rodkey T, Rojas A, Rose X, Rose R, Ruiz S, Sanders K, Savory S, Scherer S, Scott G, Shatteman S, Shen H, Shetty J, Shvartshayn A, Sisson I, Sitter C, Smajic D, Speed A, Sodergren B, Song X, Sorella R, Sosa C, Steinle M, Strong R, Sutton A, Svatek A, Taber P, Taylor C, Taylor T, Thomas R, Thomas S, Tingey A, Trejos Z, Ustani K, Valas R, Vera V, Villalana D, Waldron L, Walker S, Wang Q, Wang S, Warren S, Warren R, Wei X, White P, Williams G, Willson R, Wleczek S, Wooden H, Wortley K, Wright T, Wright P, Wu J, Yakub S, Yen C, Yoon V, Yu P, Zhang G, Zhou X, Zhao S, Dunn D, von Niederhauser A, Weiss R, Smith D, Holt R, Smith H, Weinstock G, and Gibbs R.A.
Direct Submission
Unpublished
2 (bases 1 to 227695)
Worley K.C.
Direct Submission
Submitted (13-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 227695)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (13-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 13, 2002 this sequence version replaced g1:21272182.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly 'a' contig-scaffold'. Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

```



```

----- Genome Center
Center: Baylor College of Medicine
Center code: BCX
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GNR
Center clone name: CH230-3F3
----- Summary Statistics
Assembly program: Atlas 3.0
Consensus quality: 19843 bases at least Q40
Consensus quality: 19843 bases at least Q40
Consensus quality: 20148 bases at least Q40
Estimated insert size: 204372, sum-of-coverage estimation
Quality coverage: 5x in Q40 bases; sum of coverage estimation
-----
* NOTE: Estimated insert size may differ from sequence length
  (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a working draft sequence. It currently
  consists of 4 contigs. The true order of the pieces
  is not known and their order in this sequence record is
  arbitrary. Gaps between the contigs are represented as
  runs of N, but the exact sizes of the gaps are unknown.
  This record will be updated with the finished sequence
  as soon as it is available and the accession number will
  be preserved.
-----
1. 225695: contig of 224690 bp in length
225696 225795: gap of unknown length
225796 227334: contig of 1539 bp in length
227335 227434: gap of unknown length
227435 226350: contig of 1116 bp in length
226351 228550: gap of unknown length
228551 229678: contig of 1028 bp in length
-----
FEATURES             Location/Qualifiers
     source            1..229678
                     organism="Rattus norvegicus"
                     mol_type="genomic DNA"
                     db_xref="taxon:10116"
                     clone="CH230-3F3"
                     70415..71226
                     /note="clone boundary"
                     clone_end=Sp6
                     site=ECOR1
                     end_sequence=RH3CJ891"
                     78874..80154
                     /note="wgs contig"
                     complement(229460..229463)
                     /note="clone boundary"
                     clone_end=T7
                     site=ECOR1
                     end_sequence=BH3CJ983"
                     222734..224012
                     /note="wgs end extension"
                     clone_end=T7"
                     224063..225695
                     /note="wgs end extension"
                     clone_end=T"
-----
BASE COUNT  63628 a 40553 c 39803 g 59348 t 26244 others
ORIGIN
-----
Quality March 2.7% Score 22; DB 2; Length 229678;
Best local similarity 100.0%; Pred. No. 2.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
-----
QY 705 TTAAATTTCTTGAGTACTTTATA 716
|||||
Dc 59287 TTAATTTCTTGAGTACTTTATA 59308
-----
RESULT 54
AC011176/c 244630 bp ENA linear HTG 05-SEP-2002
LOCUS Mus musculus chromosome 5 clone RP23-3L21 map 5, WORKING DRAFT
DEFINITION

```

```

SEQUENCE 29 unordered pieces.
AC011176
ACC:1176.3 GI:22726116
HTG: HTGS_PHASE1: HTGS_DRAFT; HTGS_FULLTOP.
Mus musculus (house mouse)
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
1 (bases 1 to 244620)
Birren,B., Nustbaum,C. and Lander,E.
Mus musculus chromosome 5, clone RP23-3L21
Unpublished
2 (bases 1 to 244620)
Birren,B., Linton,L., Nustbaum,C., Lander,E., Allen,N., Anderson,M.,
Balwin,J., Batta,N., Becker,V., Boguski,V., L., Bonkigaiter,B.,
Brown,A., Castle,A., Colangelo,X., Collins,S., Collymore,A.,
Cooke,P., DeRubeis,K., Dewar,K., Domino,X., Donnell,E., Doye,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galaga,S., Gardys,S., Grant,G., Hagois,B., Heaford,A., Horton,L.,
Howland,S., Johnson,R., Jones,C., Kaur,L., Karas,A., Klein,J.,
Lehoczky,J., Liu,C., Locke,K., Macdonald,P., Marques,N.,
McWan,P., McGurk,A., McKernan,K., McLaughlin,C., Melnick,J.,
Morrow,J., Nayler,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Paterson,K., Pollard,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talaas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,N., Wu,X.,
Wyman,D., Ye,W.J., Zimet,A. and Zody,M.
Direct Submission
Submitted (01-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 244620)
Birren,B., Nustbaum,C., Lander,E., Ali,A., Allen,N., Anderson,M.,
Barna,N., Bastien,V., Bloom,T., Boguski,V., L., Bonkigaiter,B.,
Carata,C., Chang,J., Chazaro,B., Chocepe,V., Collymore,A.,
Cook,A., Cooke,P., DeRubeis,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardys,S., Gird,S., Graham,L., Grand-Pierre,K., Hagois,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Karat,A.,
Karas,A., Kellis,C., Landers,T., Levine,R., Lindblad Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Melnick,J., Meneses,L., Mibova,T., Miensz,V.,
Murphy,T., Nayler,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,E., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,R., Ray-Ord,K., Rector,K., Riese,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupbach,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,C.,
Tesfaye,S., Theodore,C., Toham,K., Travers,M., Vassiliev,H.,
Vieira,R., Vo,A., Wilson,S., Wu,X., Wyman,D., Young,G., Zaitoun,J.,
Zerbek,S., Zimet,A. and Zody,M.
Direct Submission
Submitted (05-SEP-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 5, 2002 this sequence version replaced GI:6479032.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WBIR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: U3194
Center clone name: J121
----- Summary Statistics
Sequencing vector: M13; M7815; 29% of reads
Sequencing vector: Plasmid; n/a; 71% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 231229 bases at least Q40
Consensus quality: 237315 bases at least Q30
Consensus quality: 239816 bases at least Q20
Insert size: 25000; agarose-gel

```









```

/db_xref="GI:12720415"
/translation="MSKTCQSLDNGEVLVEREKLAIRAKGSEFNTFESDALAQL
RANYDAGEALKKAQEVYKAGINTFRANGKATFTLQNSGRITQLYVARNLREG
IYADVGQWGLGIGVKGTFKTKITNLTCTVTEVQLITKALRPDPIKHS-SDGE
PYRQYDQI:ANEQSRRTFI:PSKYIAGIREYFLSKGMEVE:PMQ:IPGASARFP
VTHNALVDVMY:MIAPLY:KRLVAGGFERFELNRFNEGVSRPHNPFIMUEYV
CAYAYHCLMDNTEFLLKAVDILGTIVPYGEURFDFPKPHERITMHDAILKYGAE
KILVKECLYDDEBRAVAKKIGIEVQKSWG:GSANVLESEVAEHHLIOPTFLMAHPA
ELSPARNDENPEVTHPELFIISGRIGNFSFELNABEQNSRFAQVAAKEAGIDE
AMFKDEPVTA:ERGLPFTAGEAGIOTFAMLPANAS:RDVITJFANRCK"
comp.e-ent(13647..1463)
/genes="pfib"
/feature="synonym: PM0193"
complement(13647..1463)
/genes="pfib"
/genes="pfib"
/codon_start=1
/transl_table=11
/product="Pfif"
/protein_id="AAK02274.1"
/db_xref="GI:12720411"
/translation="MKNEPRKAQALGKPRVALRGVATITHTLDGLEDVGLLELAIE
AFDEATHAEVAELHELFCQALAKLEPRYFSGQHDALCYVDLQAGSGGTFAQCMTEY
LRXYLRMAEKSGKFTLEMSQGVAC:KSAIT:KYSGEYAFGLP:ETGIRHLVKS
TDSNRRAT:SSAAFTVPEIDEDIDINPADLIDYRASGAGGQYVNTK:ESARI
THPSGIYVQCNRSCHRXDQMK:IKAK:YEM:KKXADKQAEKNSDIGNWS
QIRSYV:DCSRILK:RTGVENRTQAVLDGLDRETFASLKAGL"
4910..5593
/genes="3abc"
/feature="synonym: PM0191"
4910..5593
/genes="dstc"
/codon_start=1
/transl_table=11
/product="Dsrc"
/protein_id="AAK02275.1"
/db_xref="GI:12722413"
/translation="MKKILSLMLGMSSTFANSQVWVQLQXVGLSGVEVDSVYK
G:KRTAVTNGPIYTEDAKYLOKLYALSEKGPJDSVSSLLDRI:YANRMYVPA
KGEHVI:VFMDTSCYKVIKQ:KEYNDLSIT:YVYJAFPRGCVKTA:REMEAIT
ACQPOFALTEA:KNGPNKTXADNITKHVQCLG:QFVNGTDS:VTEKSELIGYLKP
ALLSELAQ"
5556..7377
/genes="redc"
/feature="synonym: PM0192"
5556..7377
/genes="redc"
/codon_start=1
/transl_table=11
/product="Redc"
/protein_id="AAK02276.1"
/db_xref="GI:12720413"
/translation="MKKLIQRPVPLGRVASSDFLLPLVFAPIQHFPQDLRTLAAL
LHPQLDQ:EAATV:LVQAYKQCE:VTVDFDADGAT:STALAVIALPQLOF:QVAF
IFNFEQYGLUS:AVAEQALEKQVLLMTVNGVSVGEVAFKQAGVKKVLI:TDHLP
PTLPDPAI:VNPINLSSCAFPKAGVGV:FYLMALAKRFELG:FNOQTQPNFD
ILDLVALTIARVLDONNRI:AYCGIAR:RASPCHNG:PALAEVANHAYRCLSD
LQPSALNAGRLD:SNVGVLEIAEONASAKALAEJLQGLQTKRE:FOGKLEA
LEICRLALQJL:LPF:ALYQACRHQVIGIVASRI:KHDFRPVAPAFVACGLK
SARVEGLH:RDALER:HCLYVYILKEGHVAAAGHTIQSF:PADQOAFNQ:VGM
LDGQRQGIWTDG:LESVLSL:TAEC:KEAGMGQAFPEPFDGFK:LIQRLVGE
KILKMWEPKQSGP:LDATAENVDTRYVFDLS:KLAK:AYKLD:NFRGRNRHQLLVD
Y:VFELGT"
7379..8353
/genes="PM0193"
7379..8353
/genes="PM0193"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK02277.1"
/db_xref="GI:12720414"
/translation="MTMWRIGSLFPLWVSAQALATP:SVSSYSYSDKSLP:EGROWF
SYOEPYQANPSEPT:J:QFFPDYDCRVKSSALD:LELYSQ:INPKVY:KELPIAEXA

```

```

HYSALIFVAKLEICAEQI:SALLFEFADKHKRYGCLSRFHDRL:WLQSGVNTDGF:KA
LYSVKAKAYEQAEKLEERYGVTFPPVW:DGRIYV:ASTLISDQY:FAVLDF:VSKL
XKEXQR"
8053..14742
/genes="pfis"
/feature="synonym: PM0194"
8053..14742
/genes="pfis"
/codon_start=1
/transl_table=11
/product="Pfis"
/protein_id="AAK02278.1"
/db_xref="GI:12720415"
/translation="MK:GIVAGAAQEVOMLQWQKRV:QVAST:YISGV:HGKAVA
LLOSGIGKVAANIGATIL:EMCKPDLV:NTGSAGGVATGLNVGD:VLSDEYVYHADVA
TAPYAKKOLPACPARFQSDKLVQAEKIAVQOQOVKGLICSDSRT:QGTPLAQ
IKADFPTRAVENATEIA:QVCHAFNVFFVVRALSGSGGSEMSFEEFLPUMAKQS
SAVLEK:DNJ"
complement(8834..10318)
/genes="pepa"
/feature="synonym: PM0195"
complement(8834..10318)
/genes="pepa"
/codon_start=1
/transl_table=11
/product="Pepa"
/protein_id="AAK02279.1"
/db_xref="GI:12720416"
/translation="MEYRVKSTAL:LDIKSN:IGLVNGE:SPSAOK:DELSQGVLSK
L:CSGE:KGI:QYLVLRHVPNYSABRVFVYAGKKGBEINEKFKQL:CDTINAVKAT
SAKEV:SYLSDIKIRKDDLVNINFSVETLET:SYFEQPKSKRSODVALTIVIFSA
EGDVAQAVNRHAKAIAL:GVRAAKDVANCPNPNVYLAECAYALAT:RSLITVILG
EKMDLDMNAYLAVSGSVNEAQLS:LYRNHPNDPAK:IVLVKGLTFDAGGILSK
PAECMDENKYMGGASVYGVNAIABLKLPLJNVIGVAGCENL:PDGNAYRPTLIT
MKG:ITVEVLT:DAEGRVL:CDTLTVYVFFPELVIDVATLITGACVVALGAHNSGLST
DOKAKOLEIAAQS:TDKAMRLP:GREYQOLKSNPAGL:NI:GRMGCA:TTACAF:LSN
FTERYWHULDIAGTAW:JQSNKCATGRP:VLLVQV:LINQATSK"
2897 a 2884 c 2308 g 3218 t
BASE COUNT 2897 a 2884 c 2308 g 3218 t
ORIGIN
Query Match 163 Score 21; DB 1; Length 10407;
Best Local Similarity 102.04; Pred No. 7.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CY 691 ATTTTATAAGCTGTAAATT 711
||| ||| ||| |||
Cb 5104 ATTTCATAGCTGTAAATT 5174
RESULT 56
AC017669/c
LOCUS
DEFINITION
AC017669
AC017669
VERSION
AC017669.1 G:16554328
KEYWORDS
HTG: HTGS PHASE2
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 bases 1 to 25290;
Adams, M. and Venter, J.C.
Direct Submission
Submitted (10-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10211671 by the submitter.
For more information on this record e-mail to fly@celeta.com.
• NOTE: This is a 'working draft' sequence.
• This sequence will be replaced
• by the finished sequence as soon as it is available and
the accession number will be preserved.
Location/Qualifiers

```

Tue Oct 21 11:34:30 2003

```

source
1. 25290
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
BASE COUNT      7859 a 5172 c 5396 g 6863 t
ORIGIN
Query March      1.6% Score 21; DB 2; Length 25290;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 21; Conservative 0; Mismatches 0; Index 0; Caps 0;
QY      807 AGATCTCTTTTATGGTGGCT 827
      ||||| ||||| |||||
DB      18099 AGATCTCTTTTATGGTGGCT 18079

RESULT 59
AC108010
LOCUS      AC108010      30190 bp      DNA      linear      PRI 26-APR 2003
DEFINITION Homo sapiens BAC clone RP13-489A5 from 7, complete sequence.
ACCESSION      AC108010
VERSION      AC108010.4      G1:19482353
KEYWORDS      HTG
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 30190)
AUTHORS      Sulston, J.E. and Waterston, R.
TITLE      Toward a complete human genome sequence
JOURNAL      Genome Res. 8 (11): 1097-1108 (1998)
MEDLINE      99063792
PUBMED      9847074
REFERENCE      2 (bases 1 to 30190)
AUTHORS      Kozlowski, A. and Pearman, C.
TITLE      The sequence of Homo sapiens BAC clone RP13 489A5
JOURNAL      Unpublished (2001)
REFERENCE      3 (bases 1 to 30190)
AUTHORS      Waterston, R.H.
TITLE      Direct Submission
JOURNAL      Submitted (24 JAN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE      4 (bases 1 to 30190)
AUTHORS      Waterston, R.H.
TITLE      Direct Submission
JOURNAL      Submitted (19-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE      5 (bases 1 to 30190)
AUTHORS      Waterston, R.H.
TITLE      Direct Submission
JOURNAL      Submitted (15-MAR-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE      6 (bases 1 to 30190)
AUTHORS      Waterston, R.
TITLE      Direct Submission
JOURNAL      Submitted (23-MAR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE      7 (bases 1 to 30190)
AUTHORS      Waterston, R.
TITLE      Direct Submission
JOURNAL      Submitted (26-APR-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT      On Mar 15, 2002 this sequence version replaced gi:18702412.
..... Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapins@wustl.edu
..... Summary Statistics
..... Center project name: H_PH0489A5

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., paired quality or 30x); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI chromosome 7 Mapping Project (Dr. D. Beck, Director, John D. McPherson in the Department of Human Genetics, Washington University), and the Washington University Genome Sequencing Center. For additional information about the map location of this sequence, see <http://www.genome.wustl.edu/oligo/oligo.html>, <http://www.genome.wustl.edu/oligo/oligo.html>, or <http://genome.wustl.edu/oligo/oligo.html>.

#### SOURCE INFORMATION:

The RP13-489A5 Human Female BAC Library was constructed using improved cloning techniques developed by Kazuo Ooegawa. The library was generated by Rachel Zhao in our laboratory. Construction was funded by a grant from the National Human Genome Research Institute (NHGRI, NIH contract #H5301165-04). The library was generated according to the new NHGRI/NIH "Guidelines for Human Subjects in Large-Scale DNA Sequencing". Female blood was obtained via a double blind selection protocol. Female blood DNA was isolated from one randomly chosen donor (out of 10 female donors) and partially digested with a combination of EcoRI and EcoRV restriction enzymes. The resulting fragments were ligated into the pBAC1.9 vector between the EcoRI sites for library segments 142 or either XbaI or PstI for library segments 344. Size selected DNA was cloned into the pBAC1.9 vector between the EcoRI sites for library segments 142 or the BamHI sites for library segments 344. The ligation products were transformed into DH10B electrocompetent cells (Gibco BRL Life Technologies). The library has been arrayed into 384-well microtiter dishes and also gridded onto 22x22cm nylon high density filters for screening by probe hybridization.

#### FEATURES

```

location/25290
1..30190
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone="RP13-489A5"
/clone_lib="RP13"
repeat_region      12..17
                    /pt_family="AT_1102"
repeat_region      92..101
                    /pt_family="AT_1102"
variation           129..131
                    /allele="C"
                    /allele="T"
                    /db_xref="dbSNP:2712168"
repeat_region      368..404
                    /pt_family="AT_1102"
repeat_region      416..443
                    /pt_family="12"
repeat_region      444..741
                    /pt_family="Alu"
                    /db_xref="dbSNP:2541434"
variation           742..778
                    /pt_family="12"

```









Tu Oct 21 11:34:30 2003

```

repeat_region 14556..14843
/rpt_family="AluSq"
repeat_region 15477..15583
/rpt_family="L2"
repeat_region complement(15720..16013)
/rpt_family="AluSc"
repeat_region complement(16056..16059)
/rpt_family="CAAAAn"
repeat_region 16107..16258
/rpt_family="AluSq/x"
repeat_region complement(16390..16609)
/rpt_family="AluSc"
repeat_region complement(16769..17064)
/rpt_family="AluSq"
repeat_region complement(17103..17398)
/rpt_family="AluSq"
unsure complement(17179..17226)
/note="Single-stranded terminator coverage."
17372..17513
/note="Single-stranded terminator coverage."
17435..17728
/rpt_family="AluSq"
repeat_region 17540..17591
/rpt_family="AluSc"
repeat_region 17927..18117
/rpt_family="AluSq/x"
repeat_region 18182..18357
/rpt_family="AluSq"
unsure 18234..18421
/note="Single-stranded terminator coverage."
complement(18386..18688)
/rpt_family="AluSq"
repeat_region 18781..18852
/rpt_family="AluSc"
repeat_region 18927..19333
/rpt_family="AluSc"
repeat_region 19431..19467
/rpt_family="AT rich"
repeat_region complement(21002..21040)
/rpt_family="CAAT"
repeat_region 21065..21171
/rpt_family="L2"
repeat_region complement(21189..21219)
/rpt_family="AT rich"
repeat_region 21285..21371
/rpt_family="L2"
repeat_region 21373..21509
/rpt_family="MER3"
repeat_region complement(21510..21812)
/rpt_family="AluSc"
repeat_region complement(21815..21918)
/rpt_family="MER3"
repeat_region 21934..22446
/rpt_family="L2"
repeat_region 22451..22742
/rpt_family="AluSc"
repeat_region 22763..23154
/rpt_family="L2"

```

Query Match: 2.6% Score 21; DB 9; Length 76542;  
Best Local Similarity 100.0%; Pred. No. 8.7;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1:96 GTTTTACTTTTTTTTCA 1216  
||| ||||| ||||| |||||

DB 33645 GTTTTACTTTTTTTTCA 33665

RESULT 62  
AC010444/c  
LOCUS  
DEFINITION Homo sapiens chromosome 5 clone CTD-233A10, complete sequence.  
ACCESSION AC010444

```

VERSION AC010444.5 GI:1226169
KEYWORDS
SOURCE
ORGANISM Homo sapiens Human
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eulipotyrrhota; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
1. Hases 1 to 19499
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE
2. Hases 1 to 19499
AUTHORS DOE Joint Genome Institute
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE
3. Hases 1 to 19499
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center
TITLE Direct Submission
JOURNAL Unpublished
COMMENT Submitted (19 MAY 2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Draft sequence produced by DOE Joint Genome Institute
FEATURES
FEATURES
SOURCE
BASE COUNT 27445 A 14430 C 18417 G 23414 T
ORIGIN
Query Match: 1.6% Score 21; DB 9; Length 91489;  

Best Local Similarity 100.0%; Pred. No. 8.7;  

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


QY 1:94 GTTTTACTTTTTTTTCA 1214  

||| ||||| ||||| |||||



DB 86504 GTTTTACTTTTTTTTCA 63484



RESULT 63  

AC122698  

LOCUS  

DEFINITION Homo sapiens chromosome 5 clone CTD-233A10, complete sequence.  

ACCESSION AC122698  

VERSION AC122698.2 GI:12261602  

KEYWORDS  

SOURCE Homo sapiens Human  

ORGANISM Homo sapiens  

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  

Mammalia; Eulipotyrrhota; Primates; Catarrhini; Hominoidea; Homo.  

REFERENCE
1. Hases 1 to 19499
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE
2. Hases 1 to 19499
AUTHORS DOE Joint Genome Institute
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE
3. Hases 1 to 19499
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center
TITLE Direct Submission
JOURNAL Unpublished
COMMENT Submitted (11 MAY 2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On May 31, 2002 this sequence version replaced GI:121206261.  

Draft Sequence Produced by DOE Joint Genome Institute


```



```

LTSDDQCAIDSKWTKLACQKDGIEFLKRVVVDLEDLIEANNSQVHGILVYVPV
FGMWDVSDVSDPTKIDGLSTONFENLYHNRIMDGTAKKCV: PCTPLAMVKI
IDNGLVDSKLANGELKGTVTVLNSEIVLGRPLAAMLAGAIVS12ING11:FC
SKRHGTIKSETNVTREELSKSD11:LVSPSPYKYNSD11:QDQ:TVINFAQCLNW
DESIOKSLVPT:GKVTIAVLEKRNLRLLRNQISNK
J01111138..12369,12477..14652
/!note="GeneID exon scores (in order of location ranges):
81.64, 201.54 - GSCC_ID cd_00819"
/codon_start=1
/product="similar to conserved hypothetical protein
[Schizosaccharomyces pombe]"
/protein_id="AAL92406.2"
/db_xref="GI:288282.8"
/translations="MIEDKNIERKEGTEVKTITDINISQIKDCENGINAIVTEIT
CEISTVLSQNSNNKGLSSSENSISQ:EQLSKDEL:INNEDNTNPLIITOLCN
KNEVELVEEEDNLIKOLTKKE:LGVLVDSPFNVCOS11:FPONOTVGVGDGEIE
KTTTITOMPTLSNDINLTSPS11:ONGYKKNVTSFLVAFN:SLPMTWLYLLTN
KQSISSATISTSKQISKVSSPSTISLSTPTLTSSTSNLNLIVSSNEFITP
PLSP:SPQQTITSSIT:TTTAVIKPKKNENILPKVDKN:MEJEXLYKHYKE
IFEYSKKSIOACNIDT:KSKLYFYFSLTGLSLAICPKYSLLTSSONSSTEQF
YITRESVEYFEL:QNLPHLPO:DYLDYQFSKRVES:KPIE:SH:SVNKNKNTN
SGSSNQKELQPT:KINWOTPIDKQCLARTLVKLLK:NGGSELILPHWV:IL
LSPENINYNKFKLES:DHETDPIVEYFELHSMAL:PIEKLJHKLJDLHLYES
KLPENANNKNNKLAATTNTTTTTTTTNNINNLISNN:SVGNSNHNES
DLANNNNKLNFLTDNENRFLKFNRR:QNI:ELKSKQITEREISLLKLF:SI
SLIKENPSYHLIRYSKRF:YV:GKKNVFLSTVHTIGALQUNCTNGQ:LELP
DQSLANNANNNSKDRKAKTAKIAGAATVAVLGMTGLAAPPITMMNFYAGSI
VAGT:GATGLSSATWLSVIFGACAKVSAAGMISAT:SGIKYQVHK:KSTSLRA:IG
VYSGNDGVSQDYOTKACVIRSVTDCYDIEIVEMKEVHLKOLVSEYCGT
IVQGVIRSAAN:ISQSLASALVP:SVLKAASVLDNPSLLKQSEKAGAVLAQILID
QYGNRP:LVSTSGSMRYALJELINSLTPKRYIS:IESVFIQSTIS:SCQKRW
SKILSLGRWYGTNPDLIKYVCSINCLSGILPAAQVAPIHIPSSSLVEXVD
LSNVLKSHLVEKEELICKLQIHID:NNIKYLPK:SPQYSGKVI1
complement (join 15046..16140,16198..16938,17392..17220))
/!note="GeneID exon scores (in order of location ranges):
140.61, 88.97, 8.44 - GSCC_ID cd_00822"
/codon_start=1
/product="similar to Mus musculus (Mouse), Probable
protonate hydratase (EC 4.2.1.49) (Urecaanase)
[im:dzaloneptioptionate hydrolase]"
/protein_id="AAO50896.1"
/db_xref="GI:288282.9"
/translations="MKDILSSLSKSLPEP:PEHPG:LSVTEHAPGPINKLT:NEKKL
ATKSLRFPNLSL:LANFELERLYGHLYVYEPRETHYENKAY:PIEYPAKSCA
AS:LMIMNRLDREVAQPNSELYTGGNSVFNWKAQYHLMYKLSMTQYTLNYS
GHEPGLNPNSSPRVIVNGVFNYSKADYERVAQQTQYQTAGSY:YIGPQ
C:VHGTITLLNAGKYLGVESLIGKVFVSGQSGAQT:KAAY:GAI:CI:VAEYSP
AAVKRHAQNMWMEKPLSIAQGNVYTLWEKLEKIDITVLGSDCTS:HNPNGGVY
PVQGSFESNLMDPEPTKFMQVETLRRHTTA:NALSAGMKFMVYKSFLEASR
AKADILKPTQKKEFYPSYVDI:MGDI:PSLPGCFERWCT:SGNKKOLEY:DR:AKEL
TEKLSQSGVPRVQCCQNCNHLA:SOAGEHKLVGQARILISCLLGRSELAVARK
AVAGTLPAPVY:SRDHEVSGTDAPWR:SNIKDGSQFCADMA:HNVDSDSPRGATW
VAHKGSGTGFEAINGGFLYLCQSKIQEKARIMLNWVYNG:IAERSWYKNAEN
TIKRAVELPSLKVITPNTSD:LNELNF
join(18549..18975,19366..19457)
/!note="GeneID exon scores (in order of location ranges):
55.58, 38.73 - GSCC_ID cd_00803"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAO50897.1"
/db_xref="GI:288282.0"
/translations="WRLSGILLILASAPAGINSQVLYVEEPQYCGSGSQSSIDSQ
SFPDSQSGSSQSSQSSQSSQSSQSSQSSQSSQSSQSSQSSQSSQSSQSSQSSQ
GIVNEGCVFYADRLPSIYTVSDNIVCTYNGTQYCNCKPKPIDT:VATSGRCYBAC
NGGTITQLSALVPIPTN:IVEAVITGECNNANSEVSYKSLT:CDNUNVA
QCGTFSFICNSTSSAVTIFGNDQCSQSPYSILHEPMYQCGQQRNVEYCVN"
join(19967..20333,20420..20683,20753..21046,21126..21437,
21609..21854,22108..22335,22397..24555)
/!note="GeneID exon scores (in order of location ranges):
7.80, 2.74, 13.02, 5.24, 6.33, 9.32, 145.37 - GSCC_ID
cd_00805"
/codon_start=1
/product="similar to Dictyostelium discoideum (slime

```

```

mol3. Hypothetical protein G01054 (Fragment)"
/protein_id="AAO50898.1"
/db_xref="GI:288282.1"
/translations="WKL:LLIL:FEK:OSQID:FIKSDNTLVKPKKSKSLFC
SESTYV:FEKFEKFEKFEKFEKFEKFEKFEKFEKFEKFEKFEKFEKFEKFEKFEK
SVKNS:SGKAF:STY:CEPFDVSKRSTY:PSLRTIPYKQKVLIVGKSKCS
LESITVCLQNTGFGYV:PYKPKDS:KMNENAT:FNSSSYTMTIT:VAYNSNICE
Query Match 100.00, Score 21, OR 3, Length 118927
Best Locat. Similarity 100.00, Pred No. 0
Matches 21, Conservative 0, Mismatches 0, Indels 0, Gaps 0
QY 749 CCAAGTAAGTATTTT 766
Db 4663 CCAAGTAAGTATTTT 4663
RESULT 65
HS25712376
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
JOURNAL
COMMENT
Submitted (22 Nov 2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, Hants, UK. Email: sequen-
ces@hinxton.sanger.ac.uk
Submitted (22 Nov 2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, Hants, UK. Email: sequen-
ces@hinxton.sanger.ac.uk
On May 25, 2001, this sequence version replaced gi:3204432.
During sequence assembly data is compared from overlapping clones
where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
EMBL, EMBL; SW, SwissProt; TrEMBL, TrEMBL; WPI, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/Wormpep/
This sequence
was finished as follows unless otherwise noted: all regions were
either double stranded or sequenced with an alternate chemistry or
covered by high quality data (i.e., phred quality >= 30); an
attempt was made to resolve all sequencing problems, such as
compressions and repeats; all regions were covered by at least one
plus/minus strand or more than one X-tube clone; and the assembly
was confirmed by restriction digest. This sequence was generated
from pair of bacterial clone mappings of human chromosome 22,
constructed by the Sanger Centre Chromosome 22 Mapping Group.
Further information can be found at
http://www.sanger.ac.uk/HGP/chr22
RPI-257123 is from the library RPI-1, constructed by the group of
Peter de Jong. For further details see
http://www.genet.org/bacpac/hw.html
VECTOR: pCVPAT
This sequence is the entire insert of clone RPI-257123. The true
right end of clone RPI-18661 is at 20121 in this sequence.
Location/Qualifiers
1..11456
/organism="Homo sapiens"
/mol_type="Genomic DNA"
/db_xref="RZPD:RPI1704:20257"
/db_xref="taxon:9606"
/contig="chr22"
/map="q13.1-13.2"
/clone="RPI-257123"
FEATURES
source

```



Query Match 1.64; Score 21; DB 9; Length 114846;  
 Best Loca 100.0%; Pred. No. 9;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

706 AATATTATTTTGTCTTCA 776  
 36116 AGTATTTTTTTTGTCTTCA 36096

AC129917 118759 bp DNA linear PR: 23-JAN 2003  
 Homo sapiens chromosome 17, clone RP11-279B10, complete sequence  
 AC129917  
 AC129917.6 GI:27877274  
 HTG  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo;  
 1 Dabes 1 to 118759  
 Birren,E., Nusbaum,C. and Lander,E.  
 Homo sapiens chromosome 17, clone RP11-279B10  
 Unpublished  
 2 (bases 1 to 118759)  
 Birren,E., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,  
 Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Bouckhalter,B.,  
 Camarata,C., Chang,J., Chazaro,H., Choerley,Y., Collymore,A.,  
 Cook,A., Cooke,P., DeRubeis,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Fato,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,  
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagsos,B.,  
 Horton,L., Huime,W., Iliev,I., Johnson,K., Jones,C., Kacat,A.,  
 Karatas,A., Kellis,C., Landers,T., Levine,R., Lindblad-Toh,K.,  
 Li,G., Maclean,C., MacDonald,P., Major,J., Matthews,C.,  
 McCarthy,M., Meldrum,C., Meneses,L., Minova,T., Miogawa,T.,  
 Murphy,T., Nguyen,C., Nicol,K., North,C., Norman,C.H.,  
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
 Plunkang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
 Ryan,J., Roy,A., Schauer,S., Schupack,R., Seaman,S., Severy,P.,  
 Smith,C., Spencer,B., Stange,Thomann,N., Stojanovic,N., Tallanta,S.,  
 Tesfaye,S., Theodore,J., Topkar,K., Travers,M., Vassiliev,H.,  
 Vo,A., Vo,A., Wilson,B., Wu,X., Wyman,C., Young,G., Zaitoun,J.,  
 Zerkow,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 118759)  
 Birren,E., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,  
 Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Bouckhalter,B.,  
 Camarata,C., Chang,J., Chazaro,H., Choerley,Y., Collymore,A.,  
 Cook,A., Cooke,P., DeRubeis,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Fato,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,  
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagsos,B.,  
 Horton,L., Huime,W., Iliev,I., Johnson,K., Jones,C., Kacat,A.,  
 Karatas,A., Kellis,C., Landers,T., Levine,R., Lindblad-Toh,K.,  
 Li,G., Maclean,C., MacDonald,P., Major,J., Matthews,C.,  
 McCarthy,M., Meldrum,C., Meneses,L., Minova,T., Miogawa,T.,  
 Murphy,T., Nguyen,C., Nicol,K., North,C., Norman,C.H.,  
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
 Plunkang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
 Ryan,J., Roy,A., Schauer,S., Schupack,R., Seaman,S., Severy,P.,  
 Smith,C., Spencer,B., Stange,Thomann,N., Stojanovic,N., Tallanta,S.,  
 Tesfaye,S., Theodore,J., Topkar,K., Travers,M., Vassiliev,H.,  
 Vo,A., Vo,A., Wilson,B., Wu,X., Wyman,C., Young,G., Zaitoun,J.,  
 Zerkow,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 118759)  
 Birren,E., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,  
 Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Bouckhalter,B.,  
 Camarata,C., Chang,J., Chazaro,H., Choerley,Y., Collymore,A.,  
 Cook,A., Cooke,P., DeRubeis,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Fato,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,  
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagsos,B.,  
 Horton,L., Huime,W., Iliev,I., Johnson,K., Jones,C., Kacat,A.,  
 Karatas,A., Kellis,C., Landers,T., Levine,R., Lindblad-Toh,K.,  
 Li,G., Maclean,C., MacDonald,P., Major,J., Matthews,C.,  
 McCarthy,M., Meldrum,C., Meneses,L., Minova,T., Miogawa,T.,  
 Murphy,T., Nguyen,C., Nicol,K., North,C., Norman,C.H.,  
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
 Plunkang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
 Ryan,J., Roy,A., Schauer,S., Schupack,R., Seaman,S., Severy,P.,  
 Smith,C., Spencer,B., Stange,Thomann,N., Stojanovic,N., Tallanta,S.,  
 Tesfaye,S., Theodore,J., Topkar,K., Travers,M., Vassiliev,H.,  
 Vo,A., Vo,A., Wilson,B., Wu,X., Wyman,C., Young,G., Zaitoun,J.,  
 Zerkow,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 4 (bases 1 to 118759)  
 Birren,E., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
 Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bicom,T.,  
 Boguslavskiy,L., Bouckhalter,B., Camarata,S., Chang,J., Choerley,Y.,  
 Collymore,A., Cooke,P., DeRubeis,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Fato,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,  
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagsos,B.,  
 Horton,L., Huime,W., Iliev,I., Johnson,K., Jones,C., Kacat,A.,  
 Karatas,A., Kellis,C., Landers,T., Levine,R., Lindblad-Toh,K.,  
 Li,G., Maclean,C., MacDonald,P., Major,J., Matthews,C.,  
 McCarthy,M., Meldrum,C., Meneses,L., Minova,T., Miogawa,T.,  
 Murphy,T., Nguyen,C., Nicol,K., North,C., Norman,C.H.,  
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
 Plunkang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
 Ryan,J., Roy,A., Schauer,S., Schupack,R., Seaman,S., Severy,P.,  
 Smith,C., Spencer,B., Stange,Thomann,N., Stojanovic,N., Tallanta,S.,  
 Tesfaye,S., Theodore,J., Topkar,K., Travers,M., Vassiliev,H.,  
 Vo,A., Vo,A., Wilson,B., Wu,X., Wyman,C., Young,G., Zaitoun,J.,  
 Zerkow,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (16-JAN-2003) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 4 (bases 1 to 118759)  
 Birren,E., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
 Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bicom,T.,  
 Boguslavskiy,L., Bouckhalter,B., Camarata,S., Chang,J., Choerley,Y.,  
 Collymore,A., Cooke,P., DeRubeis,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Fato,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,  
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagsos,B.,  
 Horton,L., Huime,W., Iliev,I., Johnson,K., Jones,C., Kacat,A.,  
 Karatas,A., Kellis,C., Landers,T., Levine,R., Lindblad-Toh,K.,  
 Li,G., Maclean,C., MacDonald,P., Major,J., Matthews,C.,  
 McCarthy,M., Meldrum,C., Meneses,L., Minova,T., Miogawa,T.,  
 Murphy,T., Nguyen,C., Nicol,K., North,C., Norman,C.H.,  
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
 Plunkang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
 Ryan,J., Roy,A., Schauer,S., Schupack,R., Seaman,S., Severy,P.,  
 Smith,C., Spencer,B., Stange,Thomann,N., Stojanovic,N., Tallanta,S.,  
 Tesfaye,S., Theodore,J., Topkar,K., Travers,M., Vassiliev,H.,  
 Vo,A., Vo,A., Wilson,B., Wu,X., Wyman,C., Young,G., Zaitoun,J.,  
 Zerkow,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (16-JAN-2003) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 4 (bases 1 to 118759)  
 Birren,E., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
 Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bicom,T.,  
 Boguslavskiy,L., Bouckhalter,B., Camarata,S., Chang,J., Choerley,Y.,  
 Collymore,A., Cooke,P., DeRubeis,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Fato,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,  
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagsos,B.,  
 Horton,L., Huime,W., Iliev,I., Johnson,K., Jones,C., Kacat,A.,  
 Karatas,A., Kellis,C., Landers,T., Levine,R., Lindblad-Toh,K.,  
 Li,G., Maclean,C., MacDonald,P., Major,J., Matthews,C.,  
 McCarthy,M., Meldrum,C., Meneses,L., Minova,T., Miogawa,T.,  
 Murphy,T., Nguyen,C., Nicol,K., North,C., Norman,C.H.,  
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
 Plunkang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
 Ryan,J., Roy,A., Schauer,S., Schupack,R., Seaman,S., Severy,P.,  
 Smith,C., Spencer,B., Stange,Thomann,N., Stojanovic,N., Tallanta,S.,  
 Tesfaye,S., Theodore,J., Topkar,K., Travers,M., Vassiliev,H.,  
 Vo,A., Vo,A., Wilson,B., Wu,X., Wyman,C., Young,G., Zaitoun,J.,  
 Zerkow,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (16-JAN-2003) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 4 (bases 1 to 118759)  
 Birren,E., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
 Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bicom,T.,  
 Boguslavskiy,L., Bouckhalter,B., Camarata,S., Chang,J., Choerley,Y.,  
 Collymore,A., Cooke,P., DeRubeis,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Fato,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,  
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagsos,B.,  
 Horton,L., Huime,W., Iliev,I., Johnson,K., Jones,C., Kacat,A.,  
 Karatas,A., Kellis,C., Landers,T., Levine,R., Lindblad-Toh,K.,  
 Li,G., Maclean,C., MacDonald,P., Major,J., Matthews,C.,  
 McCarthy,M., Meldrum,C., Meneses,L., Minova,T., Miogawa,T.,  
 Murphy,T., Nguyen,C., Nicol,K., North,C., Norman,C.H.,  
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
 Plunkang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
 Ryan,J., Roy,A., Schauer,S., Schupack,R., Seaman,S., Severy,P.,  
 Smith,C., Spencer,B., Stange,Thomann,N., Stojanovic,N., Tallanta,S.,  
 Tesfaye,S., Theodore,J., Topkar,K., Travers,M., Vassiliev,H.,  
 Vo,A., Vo,A., Wilson,B., Wu,X., Wyman,C., Young,G., Zaitoun,J.,  
 Zerkow,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (16-JAN-2

Q8 1:96 GGTCTACTTTTCTTTTCTGA 1246

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 121922)  
 Direct Submission  
 Submitted (08-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire, UK  
 CH02 ISA, UK. E-mail enquiries: humuery@sanger.ac.uk  
 requests: clonerequests@sanger.ac.uk  
 On Aug 2, 2000 this sequence version replaced gi:1585691.  
 During sequence assembly data is compared from overlapping clones where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMRL; Swi, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; information on the WORMPEP database can be found at <http://www.sanger.ac.uk/projects/c elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/RGP/Ctr20>  
 IMPORTANT: This sequence is not the entire insert of clone RP5-927M24 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.  
 The true left end of clone RP5-927M24 is at 1 in this sequence. The true left end of clone RP5-110CH3 is at 121823 in this sequence. The true right end of clone RP13-218F6 is at 65581 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP5-927M24 is from the library RPCL-5 constructed by the Group of Peter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>  
 VECTOR: pCYPAC2.

## FEATURES

Source  
 1..121922  
 /organism="Homo sapiens"  
 /mol\_type="Genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="20"  
 /clone="RP5-927M24"  
 /clone\_lib="RPC1-5"  
 repeat\_region  
 1..248  
 /note="AluX repeat: matches 59..306 of consensus"  
 repeat\_region  
 283..318  
 /note="AluJo repeat: matches 1..36 of consensus"  
 repeat\_region  
 402..659  
 /note="AluJo repeat: matches 43..297 of consensus"  
 repeat\_region  
 1..73..1463  
 /note="AluX repeat: matches 1..299 of consensus"  
 misc\_feature  
 complement(161..12023)  
 /note="match: GSS: EM:AAQ428420"  
 repeat\_region  
 1808..1884  
 /note="WER91C repeat: matches 63..140 of consensus"  
 repeat\_region  
 2584..2777  
 /note="MLT1A repeat: matches 179..356 of consensus"  
 misc\_feature  
 2898..3524  
 /note="CpG island"  
 /evidence=not\_experimental  
 gene  
 3023..7234  
 /gene="DJ927M24.1"  
 join(3023..3325,3779..3936,4975..5043,6649..6760,6846..67234)  
 /gene="DJ927M24.1"  
 /product="DJ927M24.1.3 (Novel protein, isoform 3)"  
 /note="match: ESTs: Em:BE747933 Em:BF663360 Em:BE883052

Em:BE277445 Em:BE277951 Em:BE730060 Em:BE395351;  
 /evidence=not\_experimental;  
 join(3023..3325,3779..3936,4243..4406,4975..5043,6549..6760,6846..67234)  
 /gene="DJ927M24.1"  
 /product="DJ927M24.1.3 (Novel protein, isoform 2)"  
 /note="match: ESTs: Em:AA773163 Em:RI3596 Em:RI1447;  
 Em:R67394 Em:AA527442 Em:791926 Em:RI4044  
 match: proteins: Sw:122398"  
 /evidence=not\_experimental;  
 3564..3569  
 /note="WR repeat: matches 108..238 of consensus"  
 join(3023..3325,3779..3936,4975..5043,6649..6735)  
 /gene="DJ927M24.1"  
 /product="DJ927M24.1.1 (Novel protein, isoform 1)"  
 /note="match: ESTs: Em:BE26725 Em:RI43444 Em:RI48422 Em:RI4244 Em:AB59426 Em:AI125783 Em:R87394 Em:AA46447 Em:AI513222 Em:AI103708 Em:AW515615 Em:RI9196 Em:RI2455 Em:AA38199 Em:AI106731 Em:AA77670 Em:AW510702 Em:RI14790 Em:AA10124 Em:AA18988 Em:RI769566 Em:RI1468 Em:RI130 Em:RI2024 Em:AW628424 Em:AA46207 Em:RI411 Em:AA11363 Em:AA18304 Em:AA47031 Em:AA19135 Em:AA4745 Em:AA47042 Em:AA52789"  
 /evidence=not\_experimental;  
 5634..5653  
 /gene="DJ927M24.1"  
 /note="match: STS: Em:HS4311T"  
 6347..6422  
 /note="2 copies 68 mer 9A conserved"  
 complement(646..6233)  
 /note="match: STS: Em:RI1579"  
 7216..7221  
 /gene="DJ927M24.1"  
 7234  
 /gene="DJ927M24.1"  
 7965..8266  
 /note="AluX repeat: matches 1..306 of consensus"  
 9033..9134  
 /note="WR repeat: matches 176..248 of consensus"  
 repeat\_region  
 9233..9491  
 /note="WR repeat: matches 2595..2519 of consensus"  
 repeat\_region  
 9549..10157  
 /note="AluY repeat: matches 1..296 of consensus"  
 repeat\_region  
 10351..11079  
 /note="AluX repeat: matches 1..311 of consensus"  
 repeat\_region  
 11382..11386  
 /note="AluB repeat: matches 1..302 of consensus"  
 repeat\_region  
 11558..11789  
 /note="FLAV A repeat: matches 1..132 of consensus"  
 misc\_feature  
 complement(11915..12413)  
 /note="match: GSS: EM:AAQ16485"  
 repeat\_region  
 12147..12337  
 /note="WR repeat: matches 61..254 of consensus"  
 repeat\_region  
 12462..12935  
 /note="WR repeat: matches 2494..2674 of consensus"  
 repeat\_region  
 12965..13052  
 /note="WR repeat: matches 2457..2750 of consensus"  
 repeat\_region  
 14870..13250  
 /note="WRB3 repeat: matches 118..324 of consensus"  
 repeat\_region  
 13251..13531  
 /note="AluX repeat: matches 1..246 of consensus"  
 repeat\_region  
 13534..13687  
 /note="WRB3 repeat: matches 1..318 of consensus"  
 repeat\_region  
 13958..14171  
 /note="WR repeat: matches 2483..2720 of consensus"  
 repeat\_region  
 14269..14384  
 /note="WR repeat: matches 23..147 of consensus"  
 repeat\_region  
 14410..14701  
 /note="AluX repeat: matches 1..291 of consensus"  
 repeat\_region  
 14732..15037  
 /note="LIMD3 repeat: matches 7374..7739 of consensus"  
 repeat\_region  
 15093..15267  
 /note="AluX repeat: matches 136..308 of consensus"





```
PT tissue, useful for breast cancer therapy
XX Claim 1a, 148; 206pp; German.
XX This invention describes novel human nucleic acid sequences from normal
CC breast tissue which have cytoskeletal activity. The nucleic acid sequences
CC can be used to produce and isolate full-length gene sequences. They can
CC be used to express proteins, which can be used as tools to find an
CC activity against breast cancer. The sequences can be used in sense or
CC antisense form. They are especially useful for medicaments for gene
CC therapy to treat breast cancer and for treating illnesses associated
CC with fat metabolism. AA233541-233616 represent expressed sequence tags
CC described in the method of the invention.
XX
SQ Sequence 107; BP; 281 A; 218 C; 255 G; 317 T; 0 other:

Query Match          92.8%; Score 1066.4; DB 2C; Length 1071;
Best Local Similarity 99.9%; Pred. No. 1.4e 262;
Matches 1061; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1  CAACAATGGCGGCTCCACCTAGGGGGTATGTTTTCGGGAGCAGCACCGCTCCGCTCAGG 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
5  CAACAATGGCGGCTCCACCTAGGGGGTATGTTTTCGGGAGCAGCACCGCTCCGCTCAGG 64
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
61  CCGCGTGGGCGCTTCGGGGCAAGCTTCGCTTCGAGGAGCAGTCCAGCGCTCTAGAC 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
65  CCGCGTGGGCGCTTCGGGGCAAGCTTCGCTTCGAGGAGCAGTCCAGCGCTCTAGAC 124
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
121  CTTCCAGCAGTACTTTGGTGGACAGCTGGAGTCATCTTCAGGCTTGCTTGCATCTC 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
125  CTTCCAGCAGTACTTTGGTGGACAGCTGGAGTCATCTTCAGGCTTGCTTGCATCTC 184
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
181  TGTGAGTCAGGACTATGTCATATGGCCAGCAGTACAGAGAAATTCGACCGGTGTGATC 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
185  TGTGAGTCAGGACTATGTCATATGGCCAGCAGTACAGAGAAATTCGACCGGTGTGATC 244
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
241  AGTGTATCCAGAGAGTTTCTGGATATTGCCAGACAGACAGAAATGTTTTCTACAAAAA 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
245  AGTGTATCCAGAGAGTTTCTGGATATTGCCAGACAGACAGAAATGTTTTCTACAAAAA 304
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
301  GATTTCAGTTATCTGTCCAGAACCCAGAGCAAGTATCAAGAGGATGTGTGAGAACTAA 360
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
305  GATTTCAGTTATCTGTCCAGAACCCAGAGCAAGTATCAAGAGGATGTGTGAGAACTAA 364
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
361  GGAATGATTCAGAGCGGAGAGTGCATAGTCAGAGGAGCTTGACAGAGCTGAGCGGATT 420
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
365  GGAATGATTCAGAGCGGAGAGTGCATAGTCAGAGGAGCTTGACAGAGCTGAGCGGATT 424
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
421  GATAGCAGTCTGGAGGACATCAAGTCGACACAAAGCCCGGACATCCCTCAGG 480
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
425  GGCAGCAGGTGCTGGAGGACATCAAGTCGACACAAAGCCCGGACATCCCTCAGG 484
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
481  GTCGCTTGGCTACCTGGAGCGGAGTCTGCCAAGATCCCTGACGCTTGAGGCAAGCT 540
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
485  GTCGCTTGGCTACCTGGAGCGGAGTCTGCCAAGATCCCTGACGCTTGAGGCAAGCT 544
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
541  GAGCAAGGCGCAGGAGGAGTGGCTTATGAGTGGGCTGATGCTGAGGTTGCGCCACAT 600
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
545  GAGCAAGGCGCAGGAGGAGTGGCTTATGAGTGGGCTGATGCTGAGGTTGCGCCACAT 604
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
601  TGTGCTGCTGGAGCTTGACATTTTGGAGAGAGCTCTTTGCCAGATATAGTATCATTTAG 660
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
605  TGTGCTGCTGGAGCTTGACATTTTGGAGAGAGCTCTTTGCCAGATATAGTATCATTTAG 664
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
661  TTTTATGCTCCATGGAATTTTCCACTATTTTATAGCTGTTAATTTCTTCAGTAC 720
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
665  TTTTATGCTCCATGGAATTTTCCACTATTTTATAGCTGTTAATTTCTTCAGTAC 724
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
721  TTTTATGCTCCATGGAATTTTCCACTATTTTATAGCTGTTAATTTCTTCAGTAC 780
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
725  TTTTATGCTCCATGGAATTTTCCACTATTTTATAGCTGTTAATTTCTTCAGTAC 784
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
781  AGTTAGACTGTGATATGATGACACAGATCTTTTTTATGGTGGCTTTGCTTGTTTAA 840
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```



```

QY 61 CCGCGCGGCGCTTCGGGCGCAAGCTTCGCTTCITCAGGCAGCTCCAGGCGCTCCTAGAC 120
DB 65 CCGCGCGGCGCTTCGGGCGCAAGCTTCGCTTCITCAGGCAGCTCCAGGCGCTCCTAGAC 124
QY 121 CTTCCAGCAGTACTTTGGTGGAGAGAGTGGAGTCACTTTTCGAGGCTTGCCTTTCATCTC 180
DB 125 CTTCCAGCAGTACTTTGGTGGAGAGAGTGGAGTCACTTTTCGAGGCTTGCCTTTCATCTC 184
QY 181 TGGTGAAGTCAAGTATGTCATGTCAGGACGATCAGGACAGAAATTCGACCGGCTGTGATC 240
DB 185 TGGTGAAGTCAAGTATGTCATGTCAGGACGATCAGGACAGAAATTCGACCGGCTGTGATC 244
QY 241 AGTGTATCCAGAAGTTCTTGATATTCAGAGCAGACAGAAATTTTCTTACAAAAA 300
DB 245 AGTGTATCCAGAAGTTCTTGATATTCAGAGCAGACAGAAATTTTCTTACAAAAA 304
QY 301 GATTGAGTATCTGTCAGAGAACAGAGTATGATCAAGAGGATGTCGAGAACTAA 360
DB 305 GATTGAGTATCTGTCAGAGAACAGAGTATGATCAAGAGGATGTCGAGAACTAA 364
QY 361 GGAAATCAATTACAGCGAAAGATGTCATGTCAGAGGACCTGCACAAAGCTGAGGCAAT 420
DB 365 GGAAATCAATTACAGCGAAAGATGTCATGTCAGAGGACCTGCACAAAGCTGAGGCAAT 424
QY 421 GGTAGCAGTCTGTCAGAGAACAGAGTATGATCAAGAGGATGTCGAGAACTAA 480
DB 425 GGTAGCAGTCTGTCAGAGAACAGAGTATGATCAAGAGGATGTCGAGAACTAA 484
QY 481 GCTGCTTGGCTACCTGGAGCAGGACCTGTCAGAGGACCTGCACAAAGCTGAGGCAAT 540
DB 485 GCTGCTTGGCTACCTGGAGCAGGACCTGTCAGAGGACCTGCACAAAGCTGAGGCAAT 544
QY 541 GAGTAAAGGCGCAGGAGGAGTGGCTATGATGAGGCTGATGTCAGGCTGAGGCAAT 600
DB 545 GAGTAAAGGCGCAGGAGGAGTGGCTATGATGAGGCTGATGTCAGGCTGAGGCAAT 604
QY 601 TCTGCTGTCGAGTACAGTTCGAGAGGACCTGTCAGAGGACCTGCACAAAGCTGAGGCAAT 660
DB 605 TCTGCTGTCGAGTACAGTTCGAGAGGACCTGTCAGAGGACCTGCACAAAGCTGAGGCAAT 664
QY 661 TTTATCTCTCCATTTGAGAAATTTCCACTATTTTATAGGCTGTAATTTTTCGATAC 720
DB 665 TTTATCTCTCCATTTGAGAAATTTCCACTATTTTATAGGCTGTAATTTTTCGATAC 724
QY 721 TTTATACATGTCGTGAGCTGGATTAACCAAGTAAGTATTTTTCGCTTTCGATAC 780
DB 725 TTTATACATGTCGTGAGCTGGATTAACCAAGTAAGTATTTTTCGCTTTCGATAC 784
QY 781 AGTTTACACTGTGAATATGATGACACAGATTTCTTTTATGCTGGCTTTGCTGTTTAA 840
DB 785 AGTTTACACTGTGAATATGATGACACAGATTTCTTTTATGCTGGCTTTGCTGTTTAA 844
QY 841 ATTTTTCATGACTTTTCA 859
DB 845 ATTTTTCATGACTTTTAA 863

```

## RESULT 4

ABA51741/c

ID ABA51741 standard; DNA; 476 bp.

XX

AC

XX

XX

DT 01-FEB-2002 (first entry)

XX

DE

KW

XX

XX

OS

XX

XX

PN

```

XX 09 AUG-2001.
XX 30-JAN-2001; 2001WD-1000163.
XX 04-FEB-2002; 2000US-1000112.
XX 26-MAY-2002; 2000US-1000156.
XX 30-JUN-2002; 2000US-1000158.
XX 03-AUG-2002; 2000US-1000166.
XX 21-SEP-2002; 2000US-1000187.
XX 27-SEP-2002; 2000US-1000189.
XX 04-OCT-2002; 2000GB-1000163.
XX (MOLE : MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, (Bios W), Rank DR)
XX WP1: 2001 481447/52.
XX Human genome derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver.
XX Claim 1: SEQ ID NO 4; c-1000 + sequence listing; English.
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX foetal liver. The present sequence is a single exon nucleic acid
XX probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WFO at ftp.wipo.int/pub/pat/patlib/pub_sequences.
XX Sequence 478 BP; 136 A; 111 C; 100 G; 131 T; 0 Other;

```

Query Match: 10.4%; Score 489; DP 22; Length 476;

Best Local Similarity: 100.0%; End: Nucleotide 110;

Matches 469; Conservat: 100.0%; Mismatches 0; Indels 0; Gaps 0;

```

QY 343 AGGATGTGTGAGAACTAGGATGAAATACAGGAAAGATGCTAGTCCAGAACT 402
DB 465 AGGATGTGTGAGAACTAGGATGAAATACAGGAAAGATGCTAGTCCAGAACT 410
QY 463 TGACAAAGCTGAGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCT 460
DB 469 TGACAAAGCTGAGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCT 468
QY 463 CCGCGGACATCTCTGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCT 462
DB 349 CCGCGGACATCTCTGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCT 460
QY 523 CACCTCTGAGAGCTGAGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCT 462
DB 289 CACCTCTGAGAGCTGAGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCT 460
QY 563 GTGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCT 462
DB 229 GTGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCT 462
QY 643 ATATGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCT 462
DB 169 ATATGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCT 460
QY 703 TGTAAATTTTTCAGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCT 462
DB 109 TGTAAATTTTTCAGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCT 460
QY 763 TTTTTCCTGCTTTTTCAGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCT 811
DB 49 TTTTTCCTGCTTTTTCAGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCTGTCAGAGGCT 1

```





```

XX MOLE-1 MOLECULAR DYNAMICS INC.
PA Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-48897/53.
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta
XX Claim 25; SEQ ID No 43; 694pp; English.
XX The present invention relates to single exon nucleic acid probes (SENPs).
XX the present sequence is one such probe. The probes are useful for
XX predicting a microarray for predicting, measuring and displaying gene
XX expression in samples derived from human placenta. The probes are useful
XX for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 478 BP; 136 A; 111 C; 100 G; 131 T; 0 other;
Query March 36.6%; Score 469; DB 22; Length 478;
Best Local Similarity 100.0%; Pred. No. 1.4e-110;
Matches 469; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
343 AGGATGTCAGAAATAGAGATGAATACAGCGGAAAGATGCACATAGCTAGAGAGCACT 402
DE ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
469 AGGATGTCAGAAATAGAGATGAATACAGCGGAAAGATGCACATAGCTAGAGAGCACT 410
CY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
403 TGACAAAGCTGAGCAATGATGACAGGCTGCTGACAGCATACATGCTGACGACAAAGC 462
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
409 TGACAAAGCTGAGCAATGATGACAGGCTGCTGACAGCATACATGCTGACGACAAAGC 350
CY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
463 CGCCGACATCCCTCAGGCTCTCTGGCTACTGTCAGCAAGATGCTGACGACAAAGC 522
DE ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
449 CGCCGACATCCCTCAGGCTCTCTGGCTACTGTCAGCAAGATGCTGACGACAAAGC 290
CY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
423 CACCTCTGAAGCGACAGCTGATGACAGGCTGCTGACAGCATACATGCTGACGACAAAGC 582
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
289 CACCTCTGAAGCGACAGCTGATGACAGGCTGCTGACAGCATACATGCTGACGACAAAGC 230
CY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
583 GTGAGCTTGGCAGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 642
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
429 GTGAGCTTGGCAGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 170
CY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
443 ATGATGAGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 702
DE ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
169 ATGATGAGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 110
CY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
703 TGTATATCTTGGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 762
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
109 TGTATATCTTGGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 50
CY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
763 TGTATATCTTGGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG

```

```

RESULT: 5
AA10005270
ID AA100052 standard; DNA; 478 BP.
XX
AC AA100052;
XX
XX (first entry)
XX
XX Probe #43 used to measure gene expression in human breast sample.
XX
XX Probe; human; breast disease; breast cancer; development disorders; ss;
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX Homo sapiens.
XX
XX NC_0157270-A2.

```

```

XX 03-AUG-2001.
XX
XX 29-JAN-2001; 2001WC-US00661.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 01-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX MOLE-1 MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-476286/51.
XX
XX Novel single exon nucleic acid probe used to measuring gene expression
XX in a human breast
XX Claim 25; SEQ ID No 43; 322pp; English.
XX
XX The present invention relates to novel single exon nucleic acid probes.
XX the present sequence is one such probe. The probes are useful for
XX measuring human gene expression in a human breast sample, where the probe
XX hybridizes at high stringency to a nucleic acid expressed in the human
XX breast. The probes are useful for predicting, diagnosing, grading,
XX staging, monitoring and prognosing diseases of the human breast.
XX particularly those diseases with polygenic aetiology. The diseases
XX include: breast cancer, disorders of development, inflammatory diseases
XX of the breast, fibrocystic changes, proliferative breast disease and
XX non-carcinoma tumours.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at http://wipo.int/pub/published/pct\_sequences.
XX
SQ Sequence 478 BP; 136 A; 111 C; 100 G; 131 T; 0 other;
Query March 36.6%; Score 469; DB 22; Length 478;
Best Local Similarity 100.0%; Pred. No. 1.4e-110;
Matches 469; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
343 AGGATGTCAGAAATAGAGATGAATACAGCGGAAAGATGCACATAGCTAGAGAGCACT 402
DE ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
469 AGGATGTCAGAAATAGAGATGAATACAGCGGAAAGATGCACATAGCTAGAGAGCACT 410
CY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
403 TGACAAAGCTGAGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 462
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
409 TGACAAAGCTGAGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 350
CY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
463 CGCCGACATCCCTCAGGCTCTCTGGCTACTGTCAGCAAGATGCTGACGACAAAGC 522
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
349 CGCCGACATCCCTCAGGCTCTCTGGCTACTGTCAGCAAGATGCTGACGACAAAGC 290
CY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
523 CACCTCTGAAGCGACAGCTGATGACAGGCTGCTGACAGCATACATGCTGACGACAAAGC 582
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
289 CACCTCTGAAGCGACAGCTGATGACAGGCTGCTGACAGCATACATGCTGACGACAAAGC 230
CY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
583 GTGAGCTTGGCAGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 642
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
229 GTGAGCTTGGCAGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 170
CY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
643 ATGATGAGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 702
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
169 ATGATGAGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 110
CY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
703 TGTATATCTTGGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 762
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
109 TGTATATCTTGGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 50
CY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
763 TGTATATCTTGGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG

```







135	CTTCAGCAGTACTTTGTGTGACGAGTTGGAGTCATCTTTCGAGGCTTCTCTTGTGATCTC	194
181	TGTTGAGTCCAGGACTATGTCAATGGCACCGCATCAGGAGAAATTCGAAACGGTGTGTGACG	240
195	TGTTGAGTCCAGGACTATGTCAATGGCACCGCATCAGGAGAAATTCGAAACGGTGTGTGACG	254
241	AGTGTATCCAGAAAGTTCTGTGATATTGCAAGACACAGCAATGTTTTTCTTACAAAAA	300
255	AGTGTATCCAGAAAGTTCTGTGATATTGCAAGACACAGCAATGTTTTTCTTACAAAAA	314
302	GATTCGAGTTATCTGTCAGAAACCGAGGCAAGTTATCAAGAGGATGTGTGTGAGACTAA	360
315	GATTCGAGTTTCTCTCAGAAACCGAGGCAAGTTATCAAGAGGATGTGTGTGAGACTAA	374
361	GGATAGATTTACAGCGGGAAGATSCACTAGTCCAGAAAGCACTTGCACAAAGCTGAGGCAAT	420
375	GGATAGATTTACAGCGGGAAGATSCACTAGTCCAGAAAGCACTTGCACAAAGCTGAGGCAAT	434
421	GGAGAGATGTTGTGGAGGACATCAAGTGTGA	451
435	GGAGAGATGTTGTGGAGGACATCAAGTGTGA	465

RESUME 13

RESOL. 13  
AAS76183  
ID AAS76183 standard; CDNA; 1941 BP.

13-FEB-2009 (first entry)

DE  
XX  
XX  
DNA  
ANC  
each  
novel  
human  
diagnostic  
probes  
#12187

Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnosis; genetic disorder; ss.

XX  
PN WC200:75367-A2.

11-000-2091

XX  
PF 30 MAY 2001 201WO-JS08631.XX  
PB  
31-MAR 2000  
2030LS-0540217

PR 23-AUG-2000; 2000US-0649267.

PA (HYSE-) HYSEO INC.

Pi Dymarc RT, Liu C, Tang Y<sup>†</sup>;

WPI; 2001-639362/73.

XX  
F-3026; 2501... 2011

P1: new isolated polynucleotide and encoded polypeptides, useful in:  
P2: diagnostics, forensics, gene mapping, identification of mutations  
P3: responsible for genetic disorders or other traits and to assess  
P4: biodiversity

XX  
PS Cia:m l: SEC ID No 12187: 103pp: 390:iss:

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (i) and (ii) are useful for treating







QY 129 AGTACTTGTGGAGGAGTTGGGATGCTTGGAGGATTTGCTTTCATCTGCTGAGT 186  
 DB 128 AGTACTTGTGGAGGAGTTGGGATGCTTGGAGGATTTGCTTTCATCTGCTGAGT 187  
 QY 159 CAGGACTATGTCATGTCAGGAGTATGAGGAGAAATTCGAACTGCTGATGATC 249  
 DB 158 CAGGACTATGTCATGTCAGGAGTATGAGGAGAAATTCGAACTGCTGATGATC 247  
 QY 249 CAGAAATTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 308  
 DB 248 CAGAAATTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 307  
 QY 309 TTATGCTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 368  
 DB 308 TTATGCTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 367  
 QY 369 TTACAGCGGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 428  
 DB 368 TTACAGCGGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 427  
 QY 409 GTGCTGAGGACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 488  
 DB 408 GTGCTGAGGACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 487  
 QY 449 GCTTACTGTCAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 548  
 DB 448 GCTTACTGTCAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 547  
 QY 549 GGCAGAGGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 608  
 DB 548 GGCAGAGGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 607  
 QY 609 GTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 668  
 DB 608 GTGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 667  
 QY 669 TCCATTTGAAATTTTCCATGATGATGATGATGATGATGATGATGATGATGATGATG 728  
 DB 668 TCCATTTGAAATTTTCCATGATGATGATGATGATGATGATGATGATGATGATGATG 727  
 QY 729 ATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 788  
 DB 728 ATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 787  
 QY 789 CTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 848  
 DB 788 CTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 847  
 QY 849 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 908  
 DB 848 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 907  
 QY 909 TAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 968  
 DB 908 TAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 967  
 QY 969 TGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1028  
 DB 968 TGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1027  
 QY 1029 AGAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1088  
 DB 1028 AGAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1087

RESULT 3  
 LOCUS BG141321  
 DEFINITION M18 HUVEC cDNA Library Homo sapiens CDNA, mRNA sequence.  
 ACCESSION BG141321  
 VERSION BG141321.1  
 KEYWORDS EST.

SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1. (bases 1 to 1442)  
 AUTHORS Nguyen, M.H. and Liu, C.H.  
 TITLE Identification of Genes in HUVECs  
 JOURNAL Unpublished  
 COMMENT Contact: Xai H. Nguyen, Canhui Liu  
 Dr. Xai Nguyen  
 University of California Los Angeles  
 5417 CHS, Surgery Oncology, UCLA, 10833 LeConte Ave. Los Angeles,  
 CA 90095, USA  
 Tel: 310 267 1963  
 Fax: 310 825 7575  
 Email: mainguyen@mednet.ucla.edu; liuch88@yahoo.com  
 PCR PRIMER1  
 FORWARD: 77  
 BACKWARD: 73  
 Insert Length: 1442 Seq Error: 0.00  
 Seq primer: T7 and T3  
 High quality sequence stop: 1442  
 PUBYA-NC Location/Qualifiers  
 1..1442  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /cell\_lines="Human Umbilical Vein Endothelial Cell"  
 /clone\_lib="HUVEC cDNA Library"  
 /notes="Vector: Lambda UNIZAP Vector; Site 1: EcoRI;  
 Site 2: XbaI;"

BASE COUNT 387 a 311 c 331 g 413 t  
 ORIGIN  
 Query Match 92.1% Score 1026.4; DB 10; Length 1442;  
 Best Local Similarity 99.9%; Pred. No. 8; se-159;  
 Matches 102; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 29 GTTCTCTGGGACGACCCGCTCCGCTCCAGCCCGCGGGCTTCCGGGCTTAAGCTTC 88  
 DB 398 GTTCTCTGGGACGACCCGCTCCGCTCCAGCCCGCGGGCTTCCGGGCTTAAGCTTC 457  
 QY 89 GCTTCTGAGGAGCTCCAGGCTCTTAGAGCTTCCAGCAGTACTTGGTGGAGAGT 149  
 DB 458 GCTTCTGAGGAGCTCCAGGCTCTTAGAGCTTCCAGCAGTACTTGGTGGAGAGT 517  
 QY 149 GTAGTCACTTTCAGGCTGCTTTCATCTCTGGTCACTCAGGACTATGCAATGGCAC 208  
 DB 518 GTAGTCACTTTCAGGCTGCTTTCATCTCTGGTCACTCAGGACTATGCAATGGCAC 577  
 QY 209 CGATCAGGAGAAATTCGAAACCGCTTTCATCAGTGTATCCAGAAATTTCTGGATATTC 268  
 DB 578 CGATCAGGAGAAATTCGAAACCGCTTTCATCAGTGTATCCAGAAATTTCTGGATATTC 637  
 QY 269 AAGACAGACAGAAATGTTTTCTTACAAAAAGATTGCAATGATCTGCTCCAGAAACCCAGA 328  
 DB 638 AAGACAGACAGAAATGTTTTCTTACAAAAAGATTGCAATGATCTGCTCCAGAAACCCAGA 597  
 QY 329 GCAAGTTATCAAGAGAGATGTGTGCAAGAACTAAGGAATGAATTAAGGAGAAAGATSCAT 388  
 DB 698 GCAAGTTATCAAGAGAGATGTGTGCAAGAACTAAGGAATGAATTAAGGAGAAAGATSCAT 757  
 QY 389 AGTCCAGAGCACTTCAGCAAGCTGAGGATTTGGCAGCAGGCTGCTGGAGGACATCAAGGT 448  
 DB 758 AGTCCAGAGCACTTCAGCAAGCTGAGGATTTGGCAGCAGGCTGCTGGAGGACATCAAGGT 817  
 QY 449 GCAGCAAAAAAGCCGCGGACATCCCTCAGGCTCTCTTGGCTACCTGGAGCAGGATC 508  
 DB 918 GCAGCAAAAAAGCCGCGGACATCCCTCAGGCTCTCTTGGCTACCTGGAGCAGGATC 877  
 QY 509 TGGTAAATCCCTGCACTCTGAAGTCAACCTGAGCAAGGCGGACAGGCTGCTGCTAT 568

```

Db      878  TGCCAAATCCCTGCACCTCTGAAGCAACGTCGAGCAAGGCGAGAGCGAGTGGCCCTAT 937
QY      569  GAGTGGGCTGATCGGAGGTTGGCCACACATCTCTCTGGGACTTGACATTTGCA 626
Db      938  GAGTGGGCTGATCGGAGGTTGGCCACACATCTCTCTGGGACTTGACATTTGCA 997
QY      629  GAACCTCTTGGCCAGATGAATGAGTTCAATTTAGTTTATGCTCCCATGAAATTTTCCA 698
Db      998  GAACCTCTTGGCCAGATGAATGAGTTCAATTTAGTTTATGCTCCCATGAAATTTTCCA 1057
QY      689  CTATTTTATAGCTTTTAAATTTCTGAGTCTTTTAAACATGCTGCTAGTTGATAA 748
Db      1058  CTATTTTATAGCTTTTAAATTTCTGAGTCTTTTAAACATGCTGCTAGTTGATAA 1117
QY      749  CCAAGTAAGTATTTTCTTTTCTTTAGCAAGTTTAGACTGTGAATATGATGACACAG 808
Db      1118  CCAAGTAAGTATTTTCTTTTCTTTAGCAAGTTTAGACTGTGAATATGATGACACAG 1177
QY      809  ATTCTTTTATAGTGGCTTGGCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 868
Db      1178  ATTCTTTTATAGTGGCTTGGCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 1237
QY      869  GTGGTGTGTTGTTAGTTTGGATCGGAGGAAAGATGATAGTAGGCTGAGATCAGGAG 928
Db      1238  GTGGTGTGTTGTTAGTTTGGATCGGAGGAAAGATGATAGTAGGCTGAGATCAGGAG 1297
QY      929  GGGAGTTTGTAGTGGCTTATGATATTAACCCCGCGTGGTGTGTAGAAAGTATGT 988
Db      1298  GGGAGTTTGTAGTGGCTTATGATATTAACCCCGCGTGGTGTGTAGAAAGTATGT 1357
QY      989  AAATTTGCTGTGTTTAAAGACTTTGACTTCTAGAGAGAGAGATCTAATACATATTT 1048
Db      1358  AAATTTGCTGTGTTTAAAGACTTTGACTTCTAGAGAGAGAGATCTAATACATATTT 1417
QY      1049  GTAATGTT 1056
Db      1418  GTAATGTT 1425

RESULT 4
BX385386 1201 bp mRNA linear EST 08-MAY-2003
LOCUS BX385386 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
DEFINITION Homo sapiens cDNA clone CS0DL010YN21 5'-PRIME, mRNA sequence.
ACCESSION BX385386
VERSION BX385386.1 GI:30449337
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 bases 1 to 1201
Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.
Full length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequenage
BP 101 91006 EVRY cedex - France
E-mail: seq@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 2813. For
more information about this cluster, see
http://www.genoscope.cns.fr/
281-bn/cluster.cgi?seq=CS0DL010CGL01&cluster=2813.f. Contact :
Peng Liang Email: fliang@lifetech.com URL:
http://full.length.invitrogen.com/ Invitrogen Corporation: 1600
Paradise Avenue Genoscope sequence ID : CS0DL010CGL01QPI.
Location/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DL010YN21"

FEATURES
source

```

```

/cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
/cell_line="RAMOS Cell Line"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT
25-NORMALIZED"
Note: "1st strand cDNA was primed with a Noll-cligolact"
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT 247 A 233 C 280 G 144 T 41 others
ORIGIN
Query Match: 95.7%; Score: 972.2; DB:13; Length:1201;
Best local Similarity: 96.4%; Pred. No.1:5e 143;
Matches:1006; Conservative:14; Mismatches:19; Indels:3; Gaps:37
QY 1 GAAGATATGAGTCTGCTATAGAGGATATGTTTCTGGACCCACCCGGTGGCTCAGG 60
Db 65 GAAGATATGAGTCTGCTATAGAGGATATGTTTCTGGACCCACCCGGTGGCTCAGG 124
QY 61 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 180
Db 105 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 184
QY 121 GTTCAACAGATATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 180
Db 185 GTTCAACAGATATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 244
QY 181 TATGATATGAGGATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 240
Db 245 TATGATATGAGGATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 304
QY 241 AGTGTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
Db 305 AGTGTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 364
QY 301 GATGGAGTATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 360
Db 365 GATGGAGTATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 424
QY 361 GATGGAGTATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 420
Db 425 GATGGAGTATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 484
QY 421 GATGGAGTATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 480
Db 485 GATGGAGTATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 544
QY 481 GATGGAGTATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540
Db 545 GATGGAGTATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 604
QY 541 GATGGAGTATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 600
Db 605 GATGGAGTATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 664
QY 601 GATGGAGTATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 660
Db 665 GATGGAGTATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 724
QY 661 GATGGAGTATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 720
Db 725 GATGGAGTATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 784
QY 721 GATGGAGTATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 779
Db 785 GATGGAGTATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 844
QY 780 GATGGAGTATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 839
Db 845 GATGGAGTATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 904
QY 840 GATGGAGTATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 899

```











QY 847 GCAGGACTTTCATCTTTAT- GTGTTGTTCTCTAGTTTATGATCCGAGGAAAGAGTA 905  
 DB 905 GCAGGACTTTCATCTTTAT- GTGTTGTTCTCTAGTTTATGATCCGAGGAAAGAGTA 964  
 QY 905 TAGTAGCTGAGAAATAGAGATGGAGTTTAGCTAGGAGCTTATGATATACCCG 965  
 DB 965 TAGTAGCTGAGAAATAGAGATGGAGTTTAGCTAGGAGCTTATGATATACCCG 1022  
 QY 905 CGGTGCTGTCTAGAAAGATATGATATGCT 997  
 DB 1023 GTGTGCTGTCTAGAAAGATATGATATGCT 1054

RESULT 10  
 LOCUS AL523437 1202 bp mRNA linear EST 22-MAY-2003  
 DEFINITION AL523437 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens  
 cDNA clone CS000045005 5'-UTR, mRNA sequence.  
 ACCESSION AL523437  
 VERSION AL523437.2 GI:11041609  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1201)  
 AUTHORS Li, W.B., Grubert, C., Jussieu, J. and Fellay, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 COMMENT On Feb 13, 2003: this sequence version replaced GI:12786910.  
 CONTACT: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 131 91006 Evry cedex - France  
 Email: seq@genoscope.cns.fr, web: www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 2833.f For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 seq-bin/cluster.cgi?seq=CS000045003p1&cluster=2833.f. Contact :  
 Feng Liang Email: fliang@life.com URL :  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Faraday Avenue, Genesee, sequence ID : CS000045003p1.  
 Location/Qualifiers  
 1..1201  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS000045005"  
 /tissue\_type="NEUROBLASTOMA COT 25-NORMALIZED"  
 /clone\_from="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI oligo (dr:  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with NotI and cloned into the NotI and EcoRV  
 sites of the pCMVSPORT 6 vector. Library was normalized."  
 314 a 226 c 255 g 336 t 64 others

PAGE COUNT 314 a 226 c 255 g 336 t 64 others  
 ORIGIN

Query Match: 57.5%; Score 866; DB 9; Length 1201;  
 Best Local Similarity 91.7%; Pred. NO. 1.8e 132;  
 Matches 928; Conservative 26; Mismatches 51; Indels 7; Gaps 4;

QY 1 CAACATGCGGGCTCCACTAGGGGGTATGTTTCTGGGCAKCCACCCGCTCCCGCTCAGG 60  
 DB 66 CAACATGCGGGCTCCACTAGGGGGTATGTTTCTGGGCAKCCACCCGCTCCCGCTCAGG 125  
 QY 61 CCGCGCGGGCTTCGGGGCCAGAGCTTCGCTTCCTCAGGACAGCTCCAGGGCTCTAGAC 120  
 DB 126 CCGCGCGGGCTTCGGGGCCAGAGCTTCGCTTCCTCAGGACAGCTCCAGGGCTCTAGAC 185  
 QY 121 CTTCCACAGCTACTTGTGTGACAGCTTGGAGTCATCTTTCGAGGCTTGGCTTGCATCTC 180  
 DB 186 CTTCCACAGCTACTTGTGTGACAGCTTGGAGTCATCTTTCGAGGCTTGGCTTGCATCTC 245

181 TGGTGTGTCAGGACTATCTCATGTCGACCC-GATCAGGAGAAATTCGAAACGGTGTGAT 239  
 246 TGGTGTGTCAGGACTATCTCATGTCGACCC-GATCAGGAGAAATTCGAAACGGTGTGAT 305  
 240 CAGTGTATCCAGAGTTTCTGTGATATTCAGAGACAGACAGATGTTTTTCTTACAAAA 299  
 306 MAGTGTATCCAGAGTTTCTGTGATATTCAGAGAGAAAGAAAGATGTTTTTCTTAAAA 365  
 300 AGATTTCAGTTATCTGTCCAGAAACAGAGCAAGTTTATCAAGAGAGTGTGTAGAACTA 359  
 366 ARATTTCAGTTATCTGTCCAGAAACAGAGCAAGTTTATCAAGAGAGTGTGTAGAACTA 425  
 360 AGAATGAAATACAGCGGAAAGATTCAGTCTCCAGAGGACCTTGACAAAGCTTAGGCAT 419  
 426 AGGAATGAATACAGCGGAAAGATTCAGTCTCCAGAGGACCTTGACAAAGCTTAGGCAT 485  
 420 TGGCAGCAGTGTCTGTGAGGAGATATACGTGTCAGACAAAAAGCCGCGACATCCCTAG 479  
 486 TGGCAGCAGTGTCTGTGAGGAGATCAACGTGTCAGACAAAAAGCCGCGACATCCCTAG 545  
 480 GGCTCCTTGGCTACTGTGAGCAGGACATCTGCCAACATCCCTGCACCTCTGAACCAAG 539  
 546 GGCTCCTTGGCTACTGTGAGCAGGACATCTGCCAACATCCCTGCACCTCTGAACCAAG 605  
 540 TGGCAAAAGCGCAGAGGAGTGGCTTATGATGGGCTGATGGCTGAGGTTGGTACACA 599  
 606 TGGCAAAAGCGCAGAGGAGTGGCTTATGATGGGCTGATGGCTGAGGTTGGTACACA 665  
 600 TTCTTCTCTGTGGACTTCAGATTTTGGAGAACTCTTGGCAGATATAGTTCATTTTA 659  
 666 TTCTTCTCTGTGGACTTCAGATTTTGGAGAACTCTTGGCAGATATAGTTCATTTTA 725  
 660 GTTTATGCTCCATTGAAANAATTTCCACTATTTTATAGCTGTTAATTTCTTGA 719  
 726 GTTTATGCTCCATTGAAANAATTTCCACTATTTTATAGCTGTTAATTTCTTGA 785  
 720 GTTTATGCTCCATTGAAANAATTTCCACTATTTTATAGCTGTTAATTTCTTGA 799  
 786 GTTTATGCTCCATTGAAANAATTTCCACTATTTTATAGCTGTTAATTTCTTGA 845  
 780 AAGTTTAGACTGTGAATGATGACACAGATTTTATAGTGGGTTTGGCTGTGTTTA 839  
 846 AAGTTTAGACTGTGAATGATGACACAGATTTTATAGTGGGTTTGGCTGTGTTTA 905  
 840 AATTTTTCAGACTTTTCATCTTTTATGTTGTTTCTGTGATGATGATGATGATGAT 899  
 906 AATTTTTCAGACTTTTCATCTTTTATGTTTCTGTGATGATGATGATGATGATGAT 964  
 900 AGAGTATAGTACCTTGAGAAATCAGAGATGGAGTTTATGCTGTAGGCTTATGATTA 959  
 965 AGAGTATAGTACCTTGAGAAATCAGAGATGGAGTTTATGCTGTAGGCTTATGATTA 1019  
 960 ACCCGCGGGTGTGTGTAGAAAAGTATGATAATTTGCTCTCTTTTAAAGACT 1011  
 1020 CCGCGCGGTGTGTGTAGAAAAGTATGATAATTTGCTCTCTTTTAAAGACT 1071

RESULT 11  
 BU158756  
 LOCUS BU158756 879 bp mRNA linear EST 04-SEP-2002  
 DEFINITION AGENCOURT 7844789 NIH\_MGC\_92 Homo sapiens cDNA clone IMAGE:6044243  
 5', mRNA sequence.  
 ACCESSION BU158756  
 VERSION BU158756.1 GI:22672666  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 879)  
 AUTHORS NIH-MGC http://rgc.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

Tue Oct 21 11:34:35 2003

JOURNAL  
COMMENT

Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgaos-remail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/ILNL at:  
<http://image.llnl.gov>  
 Plate: LLAX13286 row: 1 column: 12  
 High quality sequence stop: 757.  
 Location/Qualifiers  
     1..879  
         /organism="Homo sapiens"  
         /mol\_type="mRNA"  
         /db\_xref="taxon:9606"  
         /clone="IMAGE:604423"  
         /issue\_type="embryonal carcinoma, cell line"  
         /lab\_host="DH10B (phage resistant)"  
         /clone\_id="NIH MGC 92"  
         /note="Organ: testis; Vector: pMW-SPORT6; Site 1: Nuclei  
         Site 2: Sali; Cloned unidirectionally; Cloned dt primer 1  
         Average insert size 2 kb; Library enriched for  
         full-length clones and constructed by Life Technologies.  
         Note: this is a NIH MGC Library."  
     217 a 192 c 211 g 259 t

```

Query Match: 66.7%; Score 851.8; DS 13; Length 879;
Best Local Similarity 99.0%; Pred. KS: 1.9e-150;
Matches 870; Conservative 0; Mismatches 7; Indels 2; Gaps 1

QY 1 CAACAGTGGCGSCTCCACTAGAGGGTATGTTTCTGGGAGCCACCGGTGCCCCCTCAGG 60
   |||||
DB 1 CAACAGTGGCGSCTCCACTAGGGGTATGTTTCTGGGAGCCACCGGTGCCCCCTCAGG 60
   |||||

QY 61 GCGCGGCGGCGCTTCGGGGCGAACTTCGCTTCTTCAGCGAGCTCCAGTGGGTCTAGAC 120
   |||||
DB 61 GCGCGGCGGCGCTTCGGGGCGAACTTCGCTTCTTCAGCGAGCTCCAGTGGGTCTAGAC 120
   |||||

QY 121 CTTCAGACAGTACTTGGTGGACAGATTGGAGTCTCATCTTTCGAGGCTTGCTTGGCATCTC 180
   |||||
DB 121 CTTCAGACAGTACTTGGTGGACAGATTGGAGTCTCATCTTTCGAGGCTTGCTTGGCATCTC 180
   |||||

QY 181 TGGTGGATCAGAGACTATGTCGAATGCACCCACACAGGAAGAAATTCGAAACGGTGTGGATC 240
   |||||
DB 181 TGGTGGATCAGAGACTATGTCGAATGCACCCATCAGGAAGAAATTCGAAACGGTGTGGATC 240
   |||||

QY 241 AGTGTATCCAGAAAGTTTCGGATATTTCGAACAGACAGACAGAAATGTTTTTCTTACAAAAA 300
   |||||
DB 241 AGTGTATCCAGAAAGTTTCGGATATTTCGAACAGACAGACAGAAATGTTTTTCTTACAAAAA 300
   |||||

QY 301 GATTTCAGATTATCTCTCCAGAAACCCAGAGCAAGTATATCAAGAGAGATGTGTCCAGAACTAA 360
   |||||
DB 301 GATTTCAGTTATCTCTCTCCAGAAACCCAGAGCAAGTATATCAAGAGAGATGTGTCCAGAACTAA 360
   |||||

QY 361 GGAAATCAATTAACGGCGGAAGATGCACATAGTCTGAGAGAGCACTTCGACAAAGTCGAGGCATT 420
   |||||
DB 361 GGAAATCAATTAACGGCGGAAGATGCACATAGTCTGAGAGAGCACTTCGACAAAGTCGAGGCATT 420
   |||||

QY 421 GGCAGCAGGTGCTTGGAGGACATCAACGTCGAGTCAAGTACAAAAAGCCGCGAGCATCCCTCAGG 480
   |||||
DB 421 GGCAGCAGGTGCTTGGAGGACATCAACGTTGGAGGACATCAAAAGTCCGCGAGCATCCCTCAGG 480
   |||||

QY 481 GCTCTCTTGCCCTACCTTGGAGCAGGCATCTGGCAACATCCGCTTGCACTCTGAAGCCCAAGT 540
   |||||
DB 481 GCTCTCTTGCCCTACCTTGGAGCAGGCATCTGCAACATCCCTTGCACTCTGAAGCCCAAGT 540
   |||||

QY 541 GAGCAAAAGGCGAGAGCGAGTTGGCCCTATGAGTGGGCGTATGCGGTGAGGTTTGGCCACACAT 600
   |||||
DB 541 GAGCAAAAGGCGAGAGCGAGTTGGCCCTATGAGTGGGCGTATGCGGTGAGGTTTGGCCACACAT 600
   |||||

```

[illegible]

```

Query Match      66.3%  Score 849.2  JP 13  Length 1173:
Best Local Similarity 99.6%  Pred. No. 12-123:
Matches 85%  Conservative 0  Mismatches 3  Indels 0  Gaps 0:

QY  6  ATGCGGGGCTCCACTAGGGGGTATGTTTTTGGCGACGACACCGGTCGCCCTCAGGCGCG 65
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   64  ATGCGGGGCTCCACTAGGGGGTATGTTTTTGGCGACGACACCGGTCGCCCTCAGGCGCG 123

QY  66  CCGGGCCCTTCGGGCGCAAGCTTTCGCTTCTCAGGCGAGCTCGAGGGCGCTCCGAGACTCC 125

```







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Corrugon Ltd.

OK protein - protein search, using sw model

Run on: October 20, 2003, 20:12:15 : Search time 107 Seconds  
(without alignment)  
429.284 Million cell updates/sec

Title: US-10-029-137-2  
Performer score: 918  
Sequence: 1 MAAR:GGVFEQPRGPPQAR.....GSLAVIPLGAMNIPATLKPT 178

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Sequences: 93525 seqs, 254252604 residues

Total number of hits satisfying chosen parameters: 81525

Minimum DB seq length: 5

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Testing first 45 summaries

Database: SPRENBL231\*

1: sp\_archaea\*  
2: sp\_bacteria\*  
3: sp\_fungi\*  
4: sp\_human\*  
5: sp\_invertebrate\*  
6: sp\_mammal\*  
7: sp\_mbc\*  
8: sp\_organello\*  
9: sp\_phage\*  
10: sp\_plant\*  
11: sp\_rodent\*  
12: sp\_virus\*  
13: sp\_vertebrate\*  
14: sp\_unclassified\*  
15: sp\_virus\*  
16: sp\_bacteriophage\*  
17: sp\_archaea\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	918	100.0	178	Q9H204	Q9H204 homo sapien
2	870	94.6	178	Q9CQF2	Q9CQF2 mus musculus
3	858	94.6	178	Q92003	Q92003 mus musculus
4	847	92.3	193	Q9H275	Q9H275 homo sapien
5	248.5	27.1	189	Q9VQ29	Q9VQ29 drosophila
6	120	13.1	146	Q94GV9	Q94GV9 oryza sativ
7	119	13.0	156	Q9LFA5	Q9LFA5 mus musculus
8	106.5	11.6	1209	Q924W6	Q924W6 mus musculus
9	104.5	11.4	760	Q9S784	Q9S784 oryza sativ
10	98.5	10.7	667	Q9S124	Q9S124 oryza sativ
11	98.5	10.7	1055	Q974D1	Q974D1 homo sapien
12	98	10.7	375	Q9RAE0	Q9RAE0 caenorhabdi
13	97.5	10.6	759	Q9BTX7	Q9BTX7 mus musculus
14	97.5	10.6	821	Q9TFF1	Q9TFF1 homo sapien
15	97.5	10.6	845	Q9RPLJ	Q9RPLJ oryza sativ
16	97.5	10.6	993	Q9CQF3	Q9CQF3 mus musculus

17	97	10.6	655	4	Q9NY83	Q9NY83 homo sapien
18	97	10.6	712	4	Q9NY82	Q9NY82 homo sapien
19	97	10.6	749	4	Q14244	Q14244 homo sapien
20	96.5	10.5	1022	11	Q9AVG3	Q9AVG3 mus musculus
21	96	10.5	476	6	Q8WP20	Q8WP20 macaca fasc
22	95.5	10.4	424	11	Q8VDT5	Q8VDT5 mus musculus
23	95.5	10.4	827	10	Q8S5J7	Q8S5J7 oryza sativ
24	95	10.3	370	4	Q96PWA	Q96PWA homo sapien
25	95	10.3	479	11	Q8R2V4	Q8R2V4 mus musculus
26	95	10.3	504	4	Q8JYJ9	Q8JYJ9 homo sapien
27	95	10.3	642	11	Q8KJ98	Q8KJ98 mus musculus
28	95	10.3	654	10	Q93YU3	Q93YU3 arabidopsis
29	95	10.3	952	10	Q9SCR5	Q9SCR5 arabidopsis
30	94	10.2	656	4	Q94M50	Q94M50 homo sapien
31	93	10.1	203	10	Q9FJH2	Q9FJH2 arabidopsis
32	93	10.1	344	2	Q57369	Q57369 borrelia bu
33	92.5	10.1	481	13	Q9C9C1	Q9C9C1 brucydanio
34	92.5	10.1	993	3	Q9HF82	Q9HF82 pneumocysti
35	92	10.0	749	4	Q8TAJ5	Q8TAJ5 homo sapien
36	92	10.0	786	13	Q90658	Q90658 gallus gall
37	91	9.9	207	10	Q8GXK2	Q8GXK2 arabidopsis
38	91	9.9	296	12	Q91F87	Q91F87 chilo melle
39	90.5	9.9	250	5	Q9VXJ3	Q9VXJ3 drosophila
40	90.5	9.9	374	4	Q9BXJ5	Q9BXJ5 homo sapien
41	90.5	9.9	378	11	Q95MW7	Q95MW7 mus musculus
42	90.5	9.9	471	11	Q9D3F9	Q9D3F9 mus musculus
43	90.5	9.9	662	4	Q8TB39	Q8TB39 homo sapien
44	90.5	9.9	1114	5	Q8MT16	Q8MT16 drosophila
45	90.5	9.9	1153	5	Q8RY00	Q8RY00 drosophila

#### ALIGNMENTS

RESULT 1

Q9H204 PRELIMINARY: PRT: 178 AA.  
ID Q9H204  
AC Q9H204  
DT 01-VAR-2001 (TREMBlrel. 16, Created)  
DT 01-VAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
DR HYDROPHOBIC PROTEIN (Tumor ANGIOGENESIS marker) (FKSG20).  
GN FKSG20.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Nguyen M.H., Liu C.;  
RT "Full-length cDNA for Unknown#1."  
RI Submitted (NCV-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Nguyen M.H., Liu C.;  
RT "A Novel Gene as Tumor Angiogenesis Marker."  
RI Submitted (MAP-2001) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Wang Y.-G., Geng L.;  
RT "Cloning of FKSG20, a novel gene expressed in Adrenal gland tumor tissues."  
RI Submitted (NCV-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Strausberg E.;  
RT TISSUE=Lung;  
RI Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF21617; AAG18612.1; -;  
DR EMBL; AF358829; AAK3724.1; -;  
DR EMBL; AF319259; AAL12606.1; -;  
DR EMBL; BC011931; AAIL1936.1; -;  
KW Hypothetical protein.



[illegible]

11. —  
SEQUENCE FROM N.A.  
STRAIN BERKELEY.  
NEUJMEIN 2019606; PubMed:10711327  
Adams M.D., Celisner S.E., Holt P.A., Evans C.A., Gocayne J.D.,  
Amann E.L., Berg DE., Binkley J.S., Blair J.E., Boeckx  
Amannides F.G., Scherer S.E., Li P.W., Hinkins P.A., Geller R.P.,  
George R.A., Lewis S.E., Richards S., Ashburner W., Henderson S.N.,  
Sutton G.G., Wortman J.R., Vandeil Y.F., Zhang C., Chen L.Y.,  
Braden R.C., Rogers Y.H.S., Blair P.G., Chapple M., Pfeiffer R.D.,  
Wan K.H., Doyle C., Baxter B.A., Holt R., Adams M.D., Milos G.L.G.,  
Baird C.F., Rayani A.A., An H.H., Andrews Plannoch C., Baldwin D.,  
Bailow R.M., Rasm A.A., Baxter J.J., Bayraktaroglu L., Bouley E.M.,  
Borokov K.Y., Bencs P.V., Berni S.F., Bhandari D., Bolshakov S.,  
Borokov E., Botchan M.R., Boush J., Brockhart D., Brottier P.,  
Burtis K.C., Busam D.A., Butler H., Cadigan E., Center A., Chandra I.,  
Cherry J.M., Cawley S., Dahlke C., Davignon L.B., Davies P.,  
De Jagers C., Deicher A., Deng Z., Deyo A.D., Dow J., Elzer S.M.,  
Doushan K., Doup L.B., Edwards M., Ewing-Banks S., Finkbeiner S., Dunn P.,  
Durkin K.J., Evans-Christie C., Feltus F.A., Gabor C., Gelfand W.,  
Gibler C., Gabetlian A.E., Gill N.S., Gish W.B., Glaser K.,  
Glocke A., Gong F., Gottlieb M.H., Gu Z., Han P., Harris M.,  
Harris N.D., Harvey D., Heiman T.J., Heltzberg E., Hock C.,  
Houtin D., Houston K.A., Howland T.J., Wei M.W., Huganir C.,  
Tajalli M., Kalush F., Karpen S.H., Ke Z., Kennenbach A., Ketchum K.A.,  
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
Liu X., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin Z.,  
Liu X., Maitre B., McIntosh T.C., McLeod M.P., McPherson J.,  
Merkulov G., Moshina N.V., Mobaraki C., Morris J.J., Mostrel A.,  
Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.W., Nelson D.L.,  
Nelson D.R., Nelson K.A., Nixon K., Nusser D.B., Pacich J.V.,  
Paiszolo M., Pittman G.S., Pan S., Pollard C., Pail V., Reese X.G.,  
Reinert K., Remington K., Sanders R.L., Scheper H., Shen H.,  
Shue B.C., Siden-Kiamos I., Simpson M., Skupski X.P., Smith T.,  
Spier E.F., Spradling A.C., Stapleton M., Strong R., Sun E.,  
Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
Wang Z.Y., Wassarman D.A., Weinstein G.M., Weisenbach J.,  
Williams S.W., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
Ye J., Yeh P.-F., Zaveri J.S., Zhao X., Zhang Q., Zhao X., Zhang L.,  
Zheng X.H., Zhong F.W., Zhuo M., Zhou X., Zhu S., Zhu X., Smith H.O.,  
Zhang R.A., Myers E.W., Rubin G.M., Venter J.C.  
The genome sequence of *Drosophila melanogaster*.  
Science 287:2185-2195(2000).  
(12)  
SEQUENCE FROM N.A.

```

AC Q9LFA5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN F82_30 OR AT3G2860/F82_30.
OS Arabidopsis thaliana (Cruise-gar cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicotyledons; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1];
RP SEQUENCE FROM N.A.
RA Nyakatia G., Fartmann B., Danner D., Steir W., Holland R.,
RA Weichselgartner M., Meves H.W., Rudd S., Lencze K., Mayer K.F.X.,
RA Quetier F., Salanoubat M.
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2];
RP SEQUENCE FROM N.A.
RA EC Arabidopsis sequencing project;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [3];
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
RA Nakajima M., Enjo A., Kamiya A., Narusaka M., Chimizu P., Kawai J.,
RA Hayashizaki Y., Shinozaki K.;
RL "Arabidopsis thaliana full-length cDNA";
RT Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ142969; CAB68890.1;
DR EMBL: AK117706; CAB42357.1;
KW Hypothetical protein.
SQ SEQUENCE 156 AA; 17961 MW; 43C35226F4041346 CRC64;

Query Match: 13.0%; Score 119; DB 10; Length 156;
Best Local Similarity 18.2%; Pred. No. 0.014;
Matches 27; Conservative 45; Mismatches 62; Indels 14; Gaps 2;

QY 10 SQGPPGPPQAPGGLPGQNSILQAPGAPRPSSTVVELESSFEACFASLV ---SQE 63
DB 11 SQPSPSPPPRPFGI -----RSPETPSNQNDIEFVACVTALEAALLPCLPAPFLQ 62
QY 64 YVNGTQDEIRGTGVQCCCKPFLVARTQTEFFQKRLQSVQKPEQVKEQVGLNELNQ 123
DB 63 AIDKSPHSHQIDVHRARDQYEAANKKCLVFNGLKRETRAPRABSEKKKDIAMWERTK 122
QY 124 RKDAIVQKHLTKLARKWQVLEDINVOHK 151
DB 123 TKDELIKKHMLFOESCKLVKREIEKIR 150

RESULT 8
Q924W6
ID Q924W6 PRELIMINARY; PRT; 1209 AA.
AC Q924W6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE KIAA0298 protein.
GN KIAA0298.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1];
RP SEQUENCE FROM N.A.
RA Amid C., Harkeln T., Winterpacht A., Zabel B., Schmidt E.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ307670; CAC38114.1;
DR InterPro: IPR033649; Bbox_C.
DR InterPro: IPR031487; Bromodomain.
DR InterPro: IPR030276; GPCR Rhodopsn.
DR InterPro: IPR032965; P_rich_extensn.

QY 12 QPPSPQAPPPLPQCASLLQAPGAPRPSSTLVDELESSFE---ACFASLVQCYVNGT 68
InterPro: IPR006995; ZF_Sec23_Sec24.
InterPro: IPR000115; Znf_Bbox.
InterPro: IPR001965; Znf_PHR.
Pfam: PF00439; bromodomain; 1.
Pfam: PF00628; PHD; 1.
Pfam: PF06643; ZF-Bbox; 2.
Pfam: PF04810; ZF-Sec23_Sec24; 1.
PRINTS: PS05031; BROMODMAIN.
PROSITE: PS01117; PRICHTEXTEN.
SMART: SM03502; BPC; 1.
SMART: SM02247; BROMO; 1.
SMART: SM02242; PHD; 2.
PROSITE: PS05014; BROMOMAIN; 2; 1.
PROSITE: PS02397; G_PROTEIN_REPEAT_F1; 1; 1.
PROSITE: PS01119; ZNF_BBOX; 2.
PROSITE: PS01116; ZNF_PHR; 2; 1.
SEQUENCE 1221 AA; 14326 MW; A7AAB813A0F45D6C CRC64;

Query Match: 11.4%; Score 104.5; DB 10; Length 1221;
Best Local Similarity 21.0%; Pred. No. 1.6;
Matches 46; Conservative 40; Mismatches 62; Indels 65; Gaps 10;

QY 13 PPAHQAPLPPCLASLLQAAKAPRPSSTLVDELESSFEACFASLVQCYVNGT 60
DB 45 LPPFCHAPPPLPSCILASSCHSPACPSQNVDIKKKPEL ---- 806
QY 71 EEIRGVGQCTQKFEIDARCTEFLKRLQSLVQKPEQVIRHDSLRNLRKALVQ 130
DB 507 EE ---- NQKDELLAQ -PQLQLSTASPGQDQDTIVQIVIVEQAPVQ 604
QY 131 KHKTKLHKWQVEDINQVHKRPACIPQSSLAFLQASANTFAPLKT 174
DB 556 S ---- GQFFELQNTETPPA--PARKPALPVQKNTAPLPQ 591

RESULT 9
Q8S784
ID Q8S784 PRELIMINARY; PRT; 760 AA.
AC Q8S784;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 85.5 kDa protein.
GN OSUNBAS091C9.12.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4570;
RN [1];
RP SEQUENCE FROM N.A.
RC STRAIN=Nipponbare;
RA Spiegel L., De la Hustride M., Kirchhoff K., Preston R., Kurt K.,
RA Nascimben L., Baker C., Vill M.C., Estaviera T., Santos L., Miller B.,
RA Curtius D.W., Pallaja V., Shah R., King J., Seil M., Yang C., Dike S.,
RA Palmer L., Shaughnessy A., Badia N., McConchie W.R.;
PT "Genomic sequence for Oryza sativa, Nipponbare strain, clone
NT OSUNBAS091C9.12, complete sequence";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC094486; AAY04800.1;
DR Gramercy; C8754;
DR InterPro: IPR01124; Cytochrome_P450.
DR InterPro: IPR01917; Nitransf_2.
PROSITE: PS00599; AA_TRANSF_CLASS_2; 1.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Hypothetical protein.
SQ SEQUENCE 760 AA; 85514 MW; A584BC1B9A895CAC CRC64;

Query Match: 11.4%; Score 104.5; DB 10; Length 760;
Best Local Similarity 21.0%; Pred. No. 1.6;
Matches 46; Conservative 40; Mismatches 62; Indels 65; Gaps 10;

QY 12 QPPSPQAPPPLPQCASLLQAPGAPRPSSTLVDELESSFE---ACFASLVQCYVNGT 68

```



[illegible]



GenCore version: 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OK nucleic nucleic search, using sw model

Run on: October 20, 2003, 17:12:14 : Search time 12.7104 seconds  
(without alignments)  
4672.369 Million cell updates/sec

Title: US-10-029-137-10

Perfect score: 22

Sequence: 1 RAGCAGGAGGAGCAGCAGGAGG 22

Scoring table: IDENTITY NJF

Gapop 10.0, Gapext 1.0

Search: 2552756 seqs, 134779017 residues

Total number of hits satisfying chosen parameters: 510512

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post processing: Minimum Match 0

Maximum Match 100

Listing first 45 summaries

Database :	N Geneseq	134779017
1	/SIDSI/gcgdata/geneseq/geneseq-emb/NA1980.DAT	
2	/SIDSI/gcgdata/geneseq/geneseq-emb/NA1981.DAT	
3	/SIDSI/gcgdata/geneseq/geneseq-emb/NA1982.DAT	
4	/SIDSI/gcgdata/geneseq/geneseq-emb/NA1983.DAT	
5	/SIDSI/gcgdata/geneseq/geneseq-emb/NA1984.DAT	
6	/SIDSI/gcgdata/geneseq/geneseq-emb/NA1985.DAT	
7	/SIDSI/gcgdata/geneseq/geneseq-emb/NA1986.DAT	
8	/SIDSI/gcgdata/geneseq/geneseq-emb/NA1987.DAT	
9	/SIDSI/gcgdata/geneseq/geneseq-emb/NA1988.DAT	
10	/SIDSI/gcgdata/geneseq/geneseq-emb/NA1989.DAT	
11	/SIDSI/gcgdata/geneseq/geneseq-emb/NA1990.DAT	
12	/SIDSI/gcgdata/geneseq/geneseq-emb/NA1991.DAT	
13	/SIDSI/gcgdata/geneseq/geneseq-emb/NA1992.DAT	
14	/SIDSI/gcgdata/geneseq/geneseq-emb/NA1993.DAT	
15	/SIDSI/gcgdata/geneseq/geneseq-emb/NA1994.DAT	
16	/SIDSI/gcgdata/geneseq/geneseq-emb/NA1995.DAT	
17	/SIDSI/gcgdata/geneseq/geneseq-emb/NA1996.DAT	
18	/SIDSI/gcgdata/geneseq/geneseq-emb/NA1997.DAT	
19	/SIDSI/gcgdata/geneseq/geneseq-emb/NA1998.DAT	
20	/SIDSI/gcgdata/geneseq/geneseq-emb/NA1999.DAT	
21	/SIDSI/gcgdata/geneseq/geneseq-emb/NA2000.DAT	
22	/SIDSI/gcgdata/geneseq/geneseq-emb/NA2001.DAT	
23	/SIDSI/gcgdata/geneseq/geneseq-emb/NA2002.DAT	
24	/SIDSI/gcgdata/geneseq/geneseq-emb/NA2003.DAT	
25	/SIDSI/gcgdata/geneseq/geneseq-emb/NA2004.DAT	

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY:

Result No.	Score	Query Match	DB ID	Description
1	19.4	88.2	465	21 AAC01794
2	19.4	88.2	892	20 AAC01794
3	19.4	88.2	1071	20 AAC01794
4	19.4	88.2	10955	25 ABZ74584
5	17.8	82.9	249487	24 ABZ74584
6	17.4	79.1	24	22 AAK95733
7	16.8	76.4	155	21 AAA41727
8	16.8	76.4	549	21 AAC01794

9	16.8	76.4	1737	21 AAC01794
10	16.8	76.4	2440	21 AAC01794
11	16.8	76.4	11445	22 AAS42004
12	16.8	76.4	11445	22 AAK65020
13	16.8	76.4	11445	22 AAK97231
14	16.8	76.4	11445	22 AAK97231
15	16.8	76.4	11445	22 AAK97231
16	16.8	76.4	11445	22 AAK97231
17	16.8	76.4	11445	22 AAK97231
18	16.8	76.4	11445	22 AAK97231
19	16.8	76.4	11445	22 AAK97231
20	16.8	76.4	11445	22 AAK97231
21	16.8	76.4	11445	22 AAK97231
22	16.8	76.4	11445	22 AAK97231
23	16.8	76.4	11445	22 AAK97231
24	16.8	76.4	11445	22 AAK97231
25	16.8	76.4	11445	22 AAK97231
26	16.8	76.4	11445	22 AAK97231
27	16.8	76.4	11445	22 AAK97231
28	16.8	76.4	11445	22 AAK97231
29	16.8	76.4	11445	22 AAK97231
30	16.8	76.4	11445	22 AAK97231
31	16.8	76.4	11445	22 AAK97231
32	16.8	76.4	11445	22 AAK97231
33	16.8	76.4	11445	22 AAK97231
34	16.8	76.4	11445	22 AAK97231
35	16.8	76.4	11445	22 AAK97231
36	16.8	76.4	11445	22 AAK97231
37	16.8	76.4	11445	22 AAK97231
38	16.8	76.4	11445	22 AAK97231
39	16.8	76.4	11445	22 AAK97231
40	16.8	76.4	11445	22 AAK97231
41	16.8	76.4	11445	22 AAK97231
42	16.8	76.4	11445	22 AAK97231
43	16.8	76.4	11445	22 AAK97231
44	16.8	76.4	11445	22 AAK97231
45	16.8	76.4	11445	22 AAK97231

ALIGNMENTS

RESULT :

AAC01794  
ID AAC01794 standard; cDNA; 465 BP.

AC AAC01794:

DT 04 OCT 2000 (first entry)

DE Human secreted protein B' EST, SEQ ID NO: 1794.

XX Human; B' EST; expressed sequence tag; secreted protein; cDNA isolation;  
XX gene therapy; chromosome mapping; ss.

OS Homo sapiens.

XX E2031401.A2.

XX 06-SEP-2000.

PF 21 FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

PR (EST : GENSET).

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI: 2000-50038;/45.

XX P-PSDB; AAC01794.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for

Arabidopsis thalia  
Arabidopsis thalia  
Genomic sequence #  
Human immune/haema  
Human immune/haema  
Human immune/haema  
Human digestive sy  
Human secreted pro  
Genomic sequence #  
Human immune/haema  
Human digestive sy  
Frosophila melanog  
Frosophila melanog  
Human secreted pro  
Human secreted pro  
Bovine EST associ  
Human colon cancer  
Prostate adenocarc  
Human NCV10 coding  
Marine EST-derived  
Gene #3351 used to  
Human prostate cDN  
Human prostate-spe  
Prostate cancer th  
Frosophila melanog  
Human excretory re  
Human kidney relat  
Frosophila melanog  
Frosophila HD1 re  
Tobacco leaf antif  
Vycobacterium tube  
Human prostate exp  
Human prostate exp  
Human cancer relat  
Human cancer relat  
Human cancer relat  
Human neuroblastom  
Human neuroblastom  
Human PDZ protein

obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for diagnostic, forensic, gene therapy and chromosome mapping procedures -

Claim 1: SEQ ID 1792; 71pp + CD ROM; English.

The present sequence is one of a large number of 5' ESTs derived from mRNAs encoding secreted proteins. An ORF has been identified within the sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs derived from 30 different tissues. EST sequences usually correspond mainly to the 3' untranslated region (UTR) of the mRNA because they are often obtained from oligo-dT primed cDNA libraries. Such ESTs are not well suited for isolating cDNA sequences derived from the 5' ends of mRNAs and even in those cases where longer cDNA sequences have been obtained, the full 5' UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and chromosome mapping procedures. They are used to obtain upstream regulatory sequences and to design expression and secretion vectors.

Sequence 455 BP; 116 A; 113 C; 122 G; 106 T; 6 other;

Query Match: 89.2%; Score 19.4; DP 21; Length 465;  
 Best Local Similarity: 95.2%; Pred. No. 11;  
 Matches: 20; Conservative: 0; Mismatches: 1; Indels: 0; Gaps: 0;

QY 2 AACATGCGCGCTCCACTAGG 22  
 ||||| |||||  
 DB 17 AACATGCGCGCTCCACTAGG 37

RESULT 2  
 AAZ77452  
 ID AAZ77452 standard; cDNA; 882 BP;  
 AC  
 AAZ77452;  
 XX  
 DT 10-APR 2000 (first entry)  
 XX  
 DE Human ovarian tumor cDNA library derived EST fragment 3.  
 XX  
 KW Expressed sequence tag; EST; human; ovarian tumor; anticancer;  
 KW gene therapy; treatment; ss  
 XX  
 OS Homo sapiens.  
 XX  
 PN DE19817557 A1.  
 XX  
 PD 21-OCT 1999.  
 XX  
 PF 09-APR 1998; 98DE-1017557.  
 XX  
 PR 09-APR 1998; 98DE-1017557.  
 XX  
 PA (META-) METAGEN GES GENOMFORSCHUNG MSH.  
 XX  
 PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl B.  
 XX  
 WP1; 1999 591920/11.  
 DR P-PSDB; AAZ76513.  
 XX  
 PT New nucleic acid sequences expressed in ovarian, and some other, cancer  
 PT tissues, and derived polypeptides, for treatment of ovarian cancer and  
 PT identification of therapeutic agents -  
 XX  
 PS Claim 3; Page 140; 310pp; German.  
 XX  
 This invention describes novel nucleic acid (cDNA) sequences (A) which  
 CC have anticancer activity and are highly expressed in ovarian tumor  
 CC tissue (and some also in testis and breast cancer tissue). The products  
 CC of the invention can be used for gene therapy. (A) are used (i) for  
 CC recombinant expression of polypeptides (B) and (ii) to isolate complete  
 CC genes. (B) are used (i) to identify agents suitable for treatment of

ovarian cancer; (ii) directly for treating this form of cancer (including expression from gene therapy vectors) and (iii) for generation of specific antibodies. (A) are identified by assembling ESTs (expressed sequence tags) from a particular tissue type before comparison of the expression patterns. This allows a significantly longer fragment of the gene to be revealed, so should reduce the number of failures associated with the fact that ESTs from different libraries may represent different parts of the same unknown gene, distorting the estimated frequency of occurrence in a particular tissue. AAZ77450-277572 represent the human ovarian tumor cDNA library derived EST fragments described in the method of the invention and include the protein fragments represented in: AAZ76505-176538.

Sequence 882 BP; 216 A; 140 C; 206 G; 246 T; 0 other;

Query Match: 86.2%; Score 19.4; DP 20; Length 882;  
 Best Local Similarity: 91.3%; Pred. No. 11;  
 Matches: 20; Conservative: 0; Mismatches: 1; Indels: 0; Gaps: 0;

QY 2 AACATGCGCGCTCCACTAGG 22  
 ||||| |||||  
 DB 17 AACATGCGCGCTCCACTAGG 37

RESULT 3  
 AAZ33600  
 ID AAZ33600 standard; cDNA; 1071 BP;  
 XX  
 AC AAZ33600;  
 XX  
 DT 08-DEC 1999 (first entry)  
 XX  
 DE Human breast tumor-associated EST 60.  
 XX  
 KW Expressed sequence tag; EST; human; breast; cancer; cytostatic;  
 KW medicaments; gene therapy; treatment; fat metabolism; ss  
 XX  
 OS Homo sapiens.  
 XX  
 PN DE19813635 A1.  
 XX  
 PG 21-SEP 1999.  
 XX  
 PF 20-MAR-1998; 98DE-1013835.  
 XX  
 PR 20-MAR-1998; 98DE-1013835.  
 XX  
 PA (META-) METAGEN GES GENOMFORSCHUNG MSH.  
 XX  
 PI Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl B, Rosenthal A.  
 XX  
 WP1; 1999 526979/45.  
 DR P-PSDB; AAZ25426.  
 XX  
 XX Human nucleic acid sequences and protein products from normal breast  
 PT tissue, useful for breast cancer therapy -  
 XX  
 PS Claim 1a; 148; 206pp; German.  
 XX  
 This invention describes novel human nucleic acid sequences from normal  
 CC breast tissue which have cytostatic activity. The nucleic acid sequences  
 CC can be used to produce and isolate full length gene sequences. They can  
 CC be used to express proteins, which can be used as tools to find an  
 CC activity against breast cancer. The sequences can be used in sense or  
 CC antisense form. They are especially useful for medicaments for gene  
 CC therapy to treat breast cancer and for treating illnesses associated  
 CC with fat metabolism. AAZ33541-23361c represent expressed sequence tags  
 CC described in the method of the invention.  
 XX  
 QY Sequence 1071 BP; 281 A; 218 C; 255 G; 317 T; 0 other;  
 XX  
 Query Match: 88.2%; Score 19.4; DP 20; Length 1071;  
 Best Local Similarity: 95.2%; Pred. No. 11;



```

Matches 20; Conservative 6; Mismatches 1; Indels 0; Gaps 2;
QY 2 ACCATGGCGGCTCCACTAGG 22
    |||||
D6 7 AACATGGCGGCTCCACTAGG 27
    |||||

RESULT 4
ABZ74544
13 ABZ74544 standard; DNA; 10995 BP.
XX AC ABZ74544:
XX DT 13 MAY 2003 (first entry)
XX DE Secreted protein gene 348 genomic fragment HTXJ36, SEQ ID NO:131.
XX KW Human; secreted protein; cancer; tumor; hyperproliferative disorder;
XX KW autoimmune disorder; inflammation; angiogenic diseases; AIDS;
XX KW acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;
XX KW drug screening; chromosome identification; chromosome mapping;
XX KW cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;
XX KW anti-anemic; vulnerrary; chromosome 4p16; gene; ds.
XX OS Homo sapiens.
XX PN W020027513-A2.
XX FC 31-JUL-2002.
XX IF 26-MAR-2002; 2002MO-US09733.
XX PR 27-MAR-2001; 2001US-276657P.
XX PR 12-SEP-2001; 2001US-095082.
XX PR 12-SEP-2001; 2001US-095083.
XX PA CHINA : HUMAN GENOME SCI. INC.
XX FT Rosen CA, Robert SM.
XX WP1: 2003 04078/03.
XX PT New human secreted proteins and nucleic acids, useful for detecting or
XX PT treating cancer or other hyperproliferative disorders, autoimmune
XX PT disorders, inflammatory disorders, HIV disease, hepatitis or anemia -
XX PS Disclosure; Page 231-234; 247pp; English.
XX CC ABZ7321-ABZ7367 represent cDNAs corresponding to 40 human secreted
XX CC protein genes, and ABZ0947 ABZ0143 represent the proteins they encode.
XX CC ABZ7368-ABZ7467 represent human secreted protein polypeptide fragments. The
XX CC invention also encompasses antibodies specific for the secreted proteins,
XX CC the use of the secreted proteins in drug screening and recombination
XX CC vectors and host cells comprising a nucleic acid of the invention. The
XX CC secreted proteins are thought to be involved in biological activities
XX CC associated with cellular signaling, cellular differentiation, cell
XX CC migration, hormone activation and neurotransmitter activity. The
XX CC secreted proteins, nucleic acids encoding them, antibodies or antibody
XX CC fragments specific for the secreted proteins, and modulators of protein
XX CC activity are useful for diagnosing or treating cancers or other
XX CC hyperproliferative disorders. Additionally, the secreted proteins and
XX CC their nucleic acids may also be used in the treatment of autoimmune
XX CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS
XX CC (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote
XX CC wound healing. Nucleic acids of the invention may be used for chromosome
XX CC identification, chromosome mapping, in gene therapy, for identifying
XX CC individuals from minute biological samples, as hybridisation probes and
XX CC as molecular weight markers. The present sequence represents a human
XX CC secreted protein genomic fragment referred to in the disclosure of the
XX CC invention.
XX SQ Sequence 10995 BP; 2682 A; 2171 C; 2599 G; 3343 T; 0 other;

Query Match 88.2%; Score 19.4; DB 25; Length 10995;
Best Local Similarity 95.2%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 ACCATGGCGGCTCCACTAGG 22
    |||||
D6 7 AACATGGCGGCTCCACTAGG 27
    |||||

RESULT 5
ABN85733/c
13 ABN85733 standard; DNA; 249487 BP.
XX AC ABN85733:
XX DT 30-SEP-2002 (first entry)
XX DE Mouse genomic region containing the Itprc5 gene SEQ ID NO 3.
XX KW Mouse; Itprc5; taste; cell signalling; TC-1CS; food; pharmacological;
XX KW taste cell-specific ion channel subunit; gene; ds.
XX OS Mus sp.
XX PN W020254269-A1.
XX FC 11-JUL-2002.
XX IF 24-DEC-2001; 2001WO-US49808.
XX PR 23-DEC-2000; 2000US-259379P.
XX PR 21-DEC-2001; 2001US-0226185.
XX PA (REGC) : UNIV CALIFORNIA.
XX FT Zuker CS, Zhang Y.
XX WP1: 2002-583632/52.
XX CC Identifying modulators of taste signaling in taste cells for use in
XX CC food and pharmaceutical industries to customize and regulate taste. By
XX CC determining effect of the compound on a taste cell-specific ion channel
XX CC subunit -
XX PS Disclosure; Page 64-207; 306pp; English.
XX CC The invention relates to identifying (X1) a compound that modulates taste
XX CC signalling in taste cells, by contacting the compound with a eukaryotic
XX CC host cell or cell membrane which expresses a taste cell-specific ion
XX CC channel subunit (TC-1CS), and determining a functional effect of the
XX CC compound upon a transmembrane ion flux of a predetermined ion,
XX CC identifying a compound that modulates taste signalling in taste cells,
XX CC (X1) is useful for identifying a compound that modulates taste signalling
XX CC in taste cells, for identifying a compound that binds to a taste cell
XX CC specific ion channel subunit and for modulating taste signalling in taste
XX CC cells of a mammal, in particular a human. Modulators identified by (X1)
XX CC are used by the food and pharmaceutical industries to customize taste,
XX CC e.g. as additives to food or medicine so that the food or medicine tastes
XX CC different to the subject who ingests it. Bitter medicines can be made to
XX CC taste less bitter and sweet substance can be enhanced. The modulators are
XX CC useful for pharmacological and genetic modulation of taste signalling
XX CC pathways. The taste modulators can be directly administered to mammalian
XX CC subjects for modulation of taste in vivo. The present sequence is that of
XX CC the mouse genomic region containing the Itprc5 gene of the invention.
XX SQ Sequence 249487 BP; 64808 A; 61913 C; 61498 G; 61236 T; 32 other;

Query Match 80.9%; Score 17.8; DB 24; Length 249487;
Best Local Similarity 93.5%; Pred. No. 99;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 ACCATGGCGGCTCCACTAGG 22
    |||||

```

```

Db      217192 ACCATGGGGGCTCCACTAAGG 217372

RESULT 6
AA169927/c
ID      AA169927 standard; DNA; 244 BP.
XX
AC      AA169927;
XX
DT      17-DEC-2001 (first entry);
XX
DE      Probe e90.
XX
KW      Probe: silkworm; sericin; ss.
XX
OS      Synthetic.
XX
PN      JP2001245550-A.
XX
PD      11 SEP-2001.
XX
PF      03-MAR-2000; 2000JP 0059235
XX
PR      03-MAR-2000; 2000JP 0059239.
XX
PA      (NORU) NOKINSUISANSHO SANSHI KONGCHU.
XX
DR      WPI; 2001-610071/70.
XX
PT      A silkworm grade for producing sericin in a large amount.
XX
PS      Example 3; Page 12; 17pp; Japanese.
XX
CC      The present invention relates to a silkworm grade for producing sericin
CC      in a large amount. The silkworm is obtained by crossing a normal grade of
CC      strong high silk quantity line to a mutant grade of naked chrysalis line.
CC      The present sequence is a probe, which was used in an example from the
CC      present invention.
XX
SQ      Sequence 244 BP; 61 A; 70 C; 67 G; 46 T; 0 other;
XX
      Query Match      79.1%; Score 17.4; DB 22; Length 244;
      Best Local Similarity 94.7%; Pred. No. 89;
      Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
QY      2 ACCATGGGGGCTCCACTAG 20
      ||| ||||| |||||
DB      187 ACTATGGCGGCTCCACTAG 169

RESULT 7
AAA41737/c
ID      AAA41737 standard; cDNA; 155 BP.
XX
AC      AAA41737;
XX
DT      21 AUG-2000 (first entry);
XX
DE      Human secreted expressed sequence tag SEQ ID NO:477.
XX
KW      Human; mouse; xenopus; rat; secreted expressed sequence tag; SEST;
KW      expressed sequence tag; EST; probe; chemotactic; proliferative;
KW      immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KW      thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
KW      antiviral; antidiabetic; antiasthmatic; vulvaric; antiparkinsonian;
KW      antiulcer; osteopathic; neuroprotective; nootropic; anisporiatric;
KW      cerebroprotective; anticonvulsant; antidepressant; gene therapy;
KW      vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
KW      insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
KW      lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW      central nervous system disorder; Alzheimer's disease; stroke;
KW      Parkinson's disease; Huntington's disease; coagulation disorder;
KW      haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
XX
KW      tumour; infectious; depression; perianitis; ss.
XX
OS      Homo sapiens.
XX
PN      WO200021950-A1.
XX
PD      20-APR-2000.
XX
PF      15-OCT-1999; 9900 US44266.
XX
PR      15 OCT-1998; 9800-014435.
XX
PA      GENY 1 GENETICS INST INC.
XX
PI      Jacobs K, McCoy JW, Cavaliero ER, Collins Radio LA, Evans C;
PI      Metberg D, Treacy M;
XX
WP1: 2000-310907/20.
XX
CC      Isolated polynucleotides and encoded proteins, comprising secreted
CC      expressed sequence tag (SEST), useful for treating various disorders
CC      such as autoimmune, inflammatory, and central nervous system disorders.
XX
CC      Claim 1; Page 12; 13pp; English.
XX
CC      AAA41737 to AAA41749 represent specifically claimed secreted expressed
CC      sequence tags (SESTs), isolated from human, mouse, xenopus and rat
CC      tissue sources. The SESTs can have a range of activities depending on
CC      the tissues they were isolated from. The activities include:
CC      chemotactic; proliferative; immunomodulatory; haematopoietic;
CC      chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
CC      cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
CC      antiasthmatic; vulvaric; neuroprotective; osteopathic; neuroprotective;
CC      anticonvulsant; antiparkinsonian; antidepressant; cerebroprotective;
CC      vaccine; and antitubercular. The SESTs are useful as probes for the
CC      therapy and in vaccines. The SESTs are useful as probes for the
CC      identification and isolation of full length cDNAs and genomic DNA
CC      molecules which correspond to the SESTs. Proteins encoded by the SESTs
CC      are useful in assays for determining biological activity and raising
CC      antibodies. They may be useful for treatment of autoimmune disorders
CC      (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC      (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC      osteoporosis, osteoarthritis, central nervous system disorders
CC      (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC      disorders (thrombophilia, thrombosis), inflammatory disorders (Crohn's
CC      disease), tumours, bacterial, fungal or viral infections, depression and
CC      psoriasis. AAA41737 to AAA41749 represent linker variants which are given
CC      in the exemplification of the present invention.
XX
SQ      Sequence 155 BP; 34 A; 34 C; 45 G; 42 T; 0 other;
XX
      Query Match      76.4%; Score 15.9; DB 21; Length 155;
      Best Local Similarity 90.3%; Pred. No. 124-02;
      Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
QY      1 CACTATGCTGCTCCACTAG 20
      ||| ||||| |||||
DB      95 CACTATGCTGCTCCACTAG 96

RESULT 8
AAC43655
ID      AAC43655 standard; DNA; 149 BP.
XX
AC      AAC43655;
XX
DT      18-OCT-2000 (first entry);
XX
DE      Arabidopsis thaliana DNA fragment SEQ ID NO: 40014.
XX
KW      Hybridisation assay; genetic mapping; gene expression control;
KW      protein identification; signal transduction pathway;
KW      metabolic pathway; promoter; termination sequence; ss.

```





```
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0143625.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144644.
PR 20-JUL-1999; 99US-0144645.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0145866.
PR 02-AUG-1999; 99US-0146368.
PR 02-AUG-1999; 99US-0146369.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147264.
PR 04-AUG-1999; 99US-0147352.
PR 05-AUG-1999; 99US-0147152.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147323.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147915.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148654.
PR 16-AUG-1999; 99US-0149169.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151436.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.

PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 34-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158229.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159292.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      76.4%; Score 16.9; DB 2; Length 107;
Rest local similarity 90.0%; Pred. No. 2e-02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 CCATGGCGGCTCCACTAGGG 22
   |||||
DB 12 CCATGGCGGCTCCACTAGGG 1;

RESULT 10
AACG1802
ID AACG1802 standard; DNA: 2440 bp.
XX
XX
AC AACG1802;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 69782.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
FN EP1033405-A2.
XX
PC 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 21-MAR-1999; 99US-0125786.
PR 25-MAR-1999; 99US-0126264.
```













```
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249269.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250162.
PR 05-DEC-2000; 2000US-0251210.
PR 05-DEC-2000; 2000US-0251213.
PR 05-DEC-2000; 2000US-0251218.
PR 05-DEC-2000; 2000US-0251219.
PR 05-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251866.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251993.
PR 11-DEC-2000; 2000US-0254037.
PR 05 JAN 2001; 2001US-0254478.
XX CHINA: HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and
XX metastasis.
XX Disclosure; SEQ ID NO 4204; 3071pp - Sequence Listing; En3.1sb.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
XX amino acid sequences given in AAK52170 to AAK61921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patient's own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting
XX the nucleic acids into a host cell, and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/hematopoietic-related diseases, especially
XX cancers and cancer metastases of hematopoietic-derived cells. AAK64703
XX to AAK64704 represent human immune/hematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAK64169
XX represent sequences used in the exemplification of the present invention.
XX
XX Sequence 11445 RP; 3485 A; 1843 G; 1242 G; 3875 T; 0 other;
XX
XX Query Match. 74.4% Score 14.0, SE 24; Identity 11445;
XX Best Local Similarity 90.0%; Pred. No. Gaps: 0;
XX Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 CACCATGGCGCTCACTAG 20
XX |||||
XX 1974 CACCATGGCTACTCACTAG 1956
XX
XX RESULT 14
XX AAK52645/2
XX ID AAK52645 standard; DNA; 11445 BP.
XX AC AAK52645;
XX
XX DT 05-NOV 2001 (first entry)
XX
XX DE Human digestive system antigen genomic sequence SEQ ID NO: 4221.
XX
XX KW Human; digestive system antigen; gene therapy; cancer; appendicitis;
XX ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
XX I-gestive system disorder; Meckel's diverticulum; ds.
XX
```

```
CS Homo sapiens.
XX WO200155314 A2.
XX C2-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01324.
XX
XX 31-JAN-2000; 2000US-0129565.
XX 04-FEB-2000; 2000US-0150628.
XX 24-FEB-2000; 2000US-0154664.
XX 02-MAR-2000; 2000US-0156350.
XX 16-MAR-2000; 2000US-0159874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198423.
XX 19-MAY-2000; 2000US-0205515.
XX 27-JUN-2000; 2000US-0209457.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0215647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218295.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225457.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226688.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227209.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229511.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231143.
XX 08-SEP-2000; 2000US-0231144.
XX 08-SEP-2000; 2000US-0232080.
XX 09-SEP-2000; 2000US-0232081.
XX 12-SEP-2000; 2000US-0231969.
XX 14-SEP-2000; 2000US-0232397.
XX 14-SEP-2000; 2000US-0232396.
XX 14-SEP-2000; 2000US-0212398.
XX 14-SEP-2000; 2000US-0212399.
XX 14-SEP-2000; 2000US-0232400.
XX 14-SEP-2000; 2000US-0232401.
XX 14-SEP-2000; 2000US-0233063.
XX 14-SEP-2000; 2000US-0233064.
XX 14-SEP-2000; 2000US-0233065.
XX 21-SEP-2000; 2000US-0234223.
XX 21-SEP-2000; 2000US-0234274.
XX 25-SEP-2000; 2000US-0234937.
XX 25-SEP-2000; 2000US-0234938.
XX 26-SEP-2000; 2000US-0235484.
```



XX The invention relates to novel human genes (AB266891-AB268209) and the  
 CC encoded secreted proteins (ABP90470-ABP99872) useful for preventing,  
 CC treating or ameliorating medical conditions e.g. by protein or gene  
 CC therapy. The genes are isolated from a range of human tissues disclosed  
 CC in the specification. The nucleic acids, proteins, antibodies and  
 CC (ant)agonists are useful in the diagnosis, treatment and prevention of:  
 CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the  
 CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,  
 CC lung or prostate; (b) immune disorders e.g. Addison's disease,  
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
 CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as  
 CC myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g.  
 CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,  
 CC bacterial, fungal and parasitic infections.

XX Sequence 11445 BP; 3495 A; 1843 C; 2242 G; 3875 T; 0 other;

Query Match 76.4%; Score 16.8; DB 25; Length 1145;  
 Best Local Similarity 90.0%; Freq. No. 2, 3e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CACCATGGCGGCGCCACAG 20  
 ||||| ||||| |||||  
 Db 1974 CACCATGGCTACTCCACTAG 1955

Search completed: October 21, 2003, 06:03:06  
 Job time : 14.7104 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

CV nucleic - nucleic search, using sw model

Run on: October 20, 2003, 18:10:49 : Search time 124.791 seconds

(without alignment)  
4359.525 Million cell updates/sec

Title: us-10-029-137-10

Perfect score: 22

Sequence: 1 caccatgggagctccatagg 22

Scoring table: IDENTITY\_NUC

Gapex 10.0, Gapex 1.0

Scanned: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 455-2784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Databases: EST.\*

1: em\_estb.\*

2: em\_estm.\*

3: em\_esti.\*

4: em\_estu.\*

5: em\_estov.\*

6: em\_estp.\*

7: em\_estro.\*

8: em\_estr.\*

9: gb\_esti.\*

10: gb\_esti.\*

11: gb\_esti.\*

12: gb\_esti.\*

13: gb\_esti.\*

14: gb\_esti.\*

15: em\_estun.\*

16: em\_estom.\*

17: em\_gss\_hum.\*

18: em\_gss\_inv.\*

19: em\_gss\_phi.\*

20: em\_gss\_vit.\*

21: em\_gss\_fam.\*

22: em\_gss\_mus.\*

23: em\_gss\_mus.\*

24: em\_gss\_pro.\*

25: em\_gss\_red.\*

26: em\_gss\_phi.\*

27: em\_gss\_vit.\*

28: gb\_gss.\*

29: gb\_gss.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20.4	92.7	888	14	CD389145
2	19.4	88.2	344	9	AV686694
3	19.4	88.2	344	9	AV687554
4	19.4	88.2	344	9	AV687795

5	19.4	88.2	344	9	AV688881
6	19.4	88.2	344	9	AV689198
7	19.4	88.2	691	10	BG473678
8	19.4	88.2	728	10	BE797754
9	19.4	88.2	744	10	BG724190
10	19.4	88.2	786	10	B5445088
11	19.4	88.2	807	14	CB992984
12	19.4	88.2	831	10	BG413122
13	19.4	88.2	819	13	FU158756
14	19.4	88.2	1019	13	BX437420
15	19.4	88.2	1020	13	BX437419
16	19.4	88.2	1077	11	AF318059
17	19.4	88.2	1201	9	AL523437
18	19.4	88.2	1201	13	BX185186
19	19.4	88.2	1201	13	BX185186
20	19.4	88.2	1201	13	BX185186
21	19.4	88.2	1201	13	BX185186
22	19.4	88.2	1201	13	BX185186
23	19.4	88.2	1201	13	BX185186
24	19.4	88.2	1201	13	BX185186
25	19.4	88.2	1201	13	BX185186
26	19.4	88.2	1201	13	BX185186
27	19.4	88.2	1201	13	BX185186
28	19.4	88.2	1201	13	BX185186
29	19.4	88.2	1201	13	BX185186
30	19.4	88.2	1201	13	BX185186
31	19.4	88.2	1201	13	BX185186
32	19.4	88.2	1201	13	BX185186
33	19.4	88.2	1201	13	BX185186
34	19.4	88.2	1201	13	BX185186
35	19.4	88.2	1201	13	BX185186
36	19.4	88.2	1201	13	BX185186
37	19.4	88.2	1201	13	BX185186
38	19.4	88.2	1201	13	BX185186
39	19.4	88.2	1201	13	BX185186
40	19.4	88.2	1201	13	BX185186
41	19.4	88.2	1201	13	BX185186
42	19.4	88.2	1201	13	BX185186
43	19.4	88.2	1201	13	BX185186
44	19.4	88.2	1201	13	BX185186
45	19.4	88.2	1201	13	BX185186

## ALIGNMENTS

RESULT 1  
CD389145  
LOCUS  
AGENCOURT\_14:05122 NIH MGC 1473 Homo sapiens cDNA 5' UTR, mRNA sequence.  
CD389145  
VERSION  
EST.  
CD389145.1 31:1122774:  
EST.  
Homo sapiens (human)  
ORGANISM

CD389145  
AGENCOURT\_14:05122 NIH MGC 1473 Homo sapiens cDNA 5' UTR, mRNA sequence.  
CD389145  
VERSION  
EST.  
CD389145.1 31:1122774:  
EST.  
Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 988)  
NIH-MGC <http://imgc.ncbi.nlm.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Daniela S. Gerhard, Ph.D.  
Office of Cancer Genomics  
National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892  
Email: [cgabbs@mail.nih.gov](mailto:cgabbs@mail.nih.gov)  
Tissue Procurement: Dr. Jamie Thompson, University of WI  
cDNA Library Preparation: Gina Zastrow-Hayes  
DNA Sequencing: The I.M.A.G.E. Consortium (LNC)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Cloning distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNC at:  
<http://image.llnl.gov>

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Page 2

plate: NDKM70 row: j column: 06  
High quality sequence start: 28  
High quality sequence stop: 747

FEATURES  
SOURCE

```

1. .888
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="Trophoblast"
/lab_host="PH108 TONA"
/clone_lib="NH MCC 173"
/note="Vector: pDONR201; S
LIBR PRIMING - oligo dt: M
LIBR PROVIDER - Bradford
.98 c 215 g 232

```

BASE COUNT	223 a	198 c	215 g	252 t
ORIGIN				

```
Query Match: 92.7% Score 20.4 CR 14 Length 889
Best Local Similarity 95.5% Ref. No. 10-62
Matches 211 Conservative 0 Mismatches 11 Indels 0 Gaps 0
```

**QY** 1 CACCATGGCGGCTCCACTAGGG 22  
+ : ||| . | | | | . | | |  
**DB** 49 CAACATGGCGGTCCACTAGGG 61

RESULT 2	AV686694	344 bp	mRNA linear	EST: 167474	2002
LOCUS	AV686694				
DEFINITION	AV686694	GKC Homo sapiens cDNA clone GRCQCF12.5, mRNA sequence.			

ACCESSION AV686634  
 VERSION AV686634.1 GI:10289557  
 EST. EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE	1. Itases, J. to 344.
AUTHORS	Xu, X., Huang, G., Xu, Z., Qian, B., Zhu, Z., Yan, Q., Cai, T., Zhang, X., Xiao, H., Qu, C., Liu, F., Huang, C., Cheng, Z., Li, N., Du, S., Hu, W., Shen, K., Gu, G., Fu, G., Zhong, Y., Xu, S., Gu, W., Huang, W., Zhao, X., Hu, G., Gu, Z., Cren, Z., and Han, Z.
TITLE	Insights into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 98 (26): 15089-15094 (2001)

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (26): 15089-15094 (2001)  
MEDLINE 21625106  
PUMED 11752456

COMMENTS: Contact: Zeguang Han, Chinese National Human Genome Center at Shanghai, 455 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China  
Tel: 86-21-5880-919 (ext. 45)  
Fax: 86-21-58801522  
Email: hanzg@chgc.sh.cn  
This clone is available at CHGC in Shanghai.

```

FEATURES
SOURCE
Location/Qualifiers
1. 344
/corganism="Homo sapiens"
/mo_type="mRNA"
/db_xref="taxon:9606"
/cclone="GKGGP12"
/dev_issue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOUT"
/cclone.lib="GKC"
/note="vector: p Bluescript sk(-); Site 1: EcoRI; Site 2:
XhoI"

```

BASE COUNT:	62 a	97 c	85 g	100 t
ORIGIN				
Query Match		88.2%	Score 19.4;	DB 9;
Best Local Similarity		95.2%	Pred. No. 2.1e+02;	Length 344;

Matches	2: Conservative	3: Miscellaneous	4: Index	5: SAMP
CY	2	ACCATGGCGGCTCCATAGGG	22	
DB	5	ANCATGGCGGCTCCATAGGG	25	

[illegible]

REFERENCE  
J. Natl. Cancer Inst. 1994;86:1341.

AUTHORS  
Xia X, Huang J, Xu Z, Qian F, Zhu Z, Yan Q, Li T, Zhang X, Xia H, Gu J, Liu F, Huang W, Cheng Z, Li N, Du S, Hu W, Shen K, Lu S, Fu G, Zheng M, Xu S, Gu W, He Q, Zhu X, Hu G, Gu J, Chen Z, and Han Z.

TITLE  
Using two hepatocellular carcinomas at transposon level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver

JOURNAL  
Proc. Natl. Acad. Sci. U.S.A. 1997; 94:15094-15094 (2001)

JOURNAL: PROC. NALL. ACAD. SCI. U.S.A. 98 (26), 15089-15094 (2001)  
 MEDLINE: 21625706  
 PUBLISHED: 11752456  
 COMMENT: Contact: Zeng-Liang Huat  
 Chinese National Human Genome Center at Shanghai  
 381 USO Shouling Road, Zhangjiang Hi Tech Park, Pudong, Shanghai  
 201203, P. R. China  
 Tel: 86 21-58019191 (ex. 45)  
 Fax: 86 21-50801922  
 Email: huazhang@sc.sh.cn  
 This clone is available at CHGC in Shanghai.

```

FEATURES
source
Location/Qualifiers
1..344
/Creatinine:"Homo sapiens"
/mol_type:"RNA"
/dl_xref:"taxon:9606"
/cdname:"GKATSC2"
/tissue_type:"hepatocellular carcinoma"
/seq_start:"A341"
/lab_host:"SOLR"
/seq_pos:"1..344"

```

```

BASE COUNT      62.4      57.5      65.9      100.0
ORIGIN
Query Match     88.2%      Score 19.4      DH 9      L-arg:h 34%

```

```

QY      2 AACTATGCGGTCGATAGGG 42
        || || || || ||
Dp      5 AACTGACGCTGCTATAGG 15

```

RESULT 4	AV687795	344 bp	mRNA	linear	EST 16 JAN 1992
LOCUS	AV687795				
DEFINITION	AV687795, GK Homo sapiens cDNA clone GKC0312.5, mRNA sequence.				
ACCESSION	AV687795				
VERSION	AV687795.1	GI:10239658			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.				
REFERENCE	1 (Bases 1 to 344)				

```

AUTHOR:  Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
          Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Bu,J., Hu,W.,
          Shen,K., Ju,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
          Hu,G., Gu,J., Chen,Z., and Han,Z.
TITLE:    Insight into hepatocellular carcinogenesis at transcriptome level
          by comparing gene expression profiles of hepatocellular carcinoma
          with those of corresponding noncancerous liver
JOURNAL:  Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
MEDLINE:  21625106
PUBMED:  11752456
COMMENT:  Contact: Zeguang Han
          Chinese National Human Genome Center at Shanghai
          351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
          201203, P. R. China
          Tel: 86-21-50801919(ex.45)
          Fax: 86-21-50801922
          Email: hanzg@chgc.sh.cn
          This clone is available at CHGC in Shanghai.
FEATURES
SOURCE    Location/Qualifiers
          .. 344
          /organism="Homo sapiens"
          /mol_type="mRNA"
          /db_xref="taxon:9606"
          /clone="GKCGQ08"
          /issue_type="hepatocellular carcinoma"
          /dev_stage="Adult"
          /lab_host="SOLR"
          /clone_lib="GKC"
          /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
          XhoI"
BASE COUNT      62 a  97 c  85 g  100 t
ORIGIN
Query Match      98.2%; Score 19.4; DB 9; Length 344;
Best Local Similarity 95.2%; Pred. No. 2,1e-02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
CY      2 ACCATGGCGGCTCCACTAGG 22
          |||||
          5 MACATGGCGGCTCCACTAGG 25
RESULT 6
AV691918      392 bp      mRNA      linear      EST 16-JAN 2002
LOCUS      AV691918      GKC Homo sapiens cDNA clone GKCGSC03 5', mRNA sequence.
DEFINITION      AV691918
ACCESSION      AV691918
VERSION      AV691918.1 GI:16290744
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
          Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS      Xu,X., Huang,Q., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
          Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Bu,J., Hu,W.,
          Shen,K., Ju,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
          Hu,G., Gu,J., Chen,Z., and Han,Z.
TITLE      Insight into hepatocellular carcinogenesis at transcriptome level
          by comparing gene expression profiles of hepatocellular carcinoma
          with those of corresponding noncancerous liver
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
MEDLINE      21625106
PUBMED      11752456
COMMENT      Contact: Zeguang Han
          Chinese National Human Genome Center at Shanghai
          351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
          201203, P. R. China
          Tel: 86-21-50801919(ex.45)
          Fax: 86-21-50801922
          Email: hanzg@chgc.sh.cn
          This clone is available at CHGC in Shanghai.
FEATURES
SOURCE    Location/Qualifiers
          .. 392
          /organism="Homo sapiens"
          /mol_type="mRNA"
          /db_xref="taxon:9606"
          /clone="GKCGSC03"
          /issue_type="hepatocellular carcinoma"
          /dev_stage="Adult"
          /lab_host="SOLR"
          /clone_lib="GKC"
          /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
          XhoI"
BASE COUNT      73 a  104 c  112 g  103 t
ORIGIN
Query Match      88.2%; Score 19.4; DB 9; Length 392;
Best Local Similarity 95.2%; Pred. No. 2,2e-02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
CY      2 ACCATGGCGGCTCCACTAGG 22
          |||||
          5 MACATGGCGGCTCCACTAGG 25

```

```

source
.. 344
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GKCGQ08"
/issue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="GKC"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT      62 a  97 c  85 g  100 t
ORIGIN
Query Match      98.2%; Score 19.4; DB 9; Length 344;
Best Local Similarity 95.2%; Pred. No. 2,1e-02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
CY      2 ACCATGGCGGCTCCACTAGG 22
          |||||
          5 MACATGGCGGCTCCACTAGG 25
RESULT 6
AV691918      392 bp      mRNA      linear      EST 16-JAN 2002
LOCUS      AV691918      GKC Homo sapiens cDNA clone GKCGSC03 5', mRNA sequence.
DEFINITION      AV691918
ACCESSION      AV691918
VERSION      AV691918.1 GI:16290744
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
          Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS      Xu,X., Huang,Q., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
          Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Bu,J., Hu,W.,
          Shen,K., Ju,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
          Hu,G., Gu,J., Chen,Z., and Han,Z.
TITLE      Insight into hepatocellular carcinogenesis at transcriptome level
          by comparing gene expression profiles of hepatocellular carcinoma
          with those of corresponding noncancerous liver
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
MEDLINE      21625106
PUBMED      11752456
COMMENT      Contact: Zeguang Han
          Chinese National Human Genome Center at Shanghai
          351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
          201203, P. R. China
          Tel: 86-21-50801919(ex.45)
          Fax: 86-21-50801922
          Email: hanzg@chgc.sh.cn
          This clone is available at CHGC in Shanghai.
FEATURES
SOURCE    Location/Qualifiers
          .. 392
          /organism="Homo sapiens"
          /mol_type="mRNA"
          /db_xref="taxon:9606"
          /clone="GKCGSC03"
          /issue_type="hepatocellular carcinoma"
          /dev_stage="Adult"
          /lab_host="SOLR"
          /clone_lib="GKC"
          /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
          XhoI"
BASE COUNT      73 a  104 c  112 g  103 t
ORIGIN
Query Match      88.2%; Score 19.4; DB 9; Length 392;
Best Local Similarity 95.2%; Pred. No. 2,2e-02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
CY      2 ACCATGGCGGCTCCACTAGG 22
          |||||
          5 MACATGGCGGCTCCACTAGG 25

```



DB 5 AACATGCGGCTCCACTAGG 25  
 ||||| | |||| | ||||  
 TITLE JOURNAL

RESULT 7  
 BG473678  
 LOCUS 602516846F1 NIH\_MGC\_16 Homo sapiens cDNA clone IMAGE:4647683 5'  
 DEFINITION mRNA sequence.  
 ACCESSION BG473678  
 VERSION BG473678.1 GI:13405955  
 KEYWORDS EST  
 SOURCE Homo sapiens (human)  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 693)  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgaps@mail.nih.gov](mailto:cgaps@mail.nih.gov)  
 Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)  
 cDNA sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: [image.llnl.gov](http://image.llnl.gov)  
 Plate: LLM1424 row: X column: 12  
 High quality sequence stop: 489.  
 Location/Qualifiers  
 1..693

## FEATURES

Location/Qualifiers  
 1..693  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4647683"  
 /tissue\_type="retinoblastoma"  
 /clone\_libs="NIH\_MGC\_16"  
 /note="Organ: eye; Vector: pOTS7; Site: 1; XhoI; Site 2: EcoRI; cDNA made by oligo 3' printing. Directionally cloned into EcoR/XhoI sites using the following 5' adaptor: GGCACGAGCG. Library constructed by Ling Hong at the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library."  
 174..173 C 180 G 166 T

BASE COUNT 174 A 173 C 180 G 166 T  
 ORIGIN  
 Query Match 88.2% Score 13.4; DE 10; Length 693;  
 Best Local Similarity 95.2% Pred. No. 2.6e-02;  
 Mismatches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AACATGCGGCTCCACTAGG 22  
 ||||| | |||| | ||||  
 DB 5 AACATGCGGCTCCACTAGG 25

RESULT 8  
 BE797754  
 LOCUS 601521366F1 NIH\_MGC\_7 Homo sapiens cDNA clone IMAGE:3936713 5'  
 DEFINITION mRNA sequence.  
 ACCESSION BE797754  
 VERSION BE797754.1 GI:10218952  
 KEYWORDS EST  
 SOURCE Homo sapiens (human)  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 728)  
 NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgaps@mail.nih.gov](mailto:cgaps@mail.nih.gov)  
 Tissue Procurement: PCID/ETP  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)  
 cDNA sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: [image.llnl.gov](http://image.llnl.gov)  
 Plate: LLM711 row: C column: 14  
 High quality sequence start: 26  
 High quality sequence stop: 723.  
 Location/Qualifiers  
 1..724

## FEATURES

Location/Qualifiers  
 1..724  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4647683"  
 /tissue\_type="retinoblastoma"  
 /clone\_libs="NIH\_MGC\_16"  
 /note="Organ: eye; Vector: pOTS7; Site: 1; XhoI; Site 2: EcoRI; cDNA made by oligo 3' printing. Directionally cloned into EcoR/XhoI sites using the following 5' adaptor: GGCACGAGCG. Library constructed by Ling Hong at the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC Library."  
 174..173 C 180 G 166 T

## BASE COUNT

ORIGIN  
 Query Match 88.2% Score 19.4; DE 10; Length 724;  
 Best Local Similarity 95.2% Pred. No. 2.6e-02;  
 Mismatches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AACATGCGGCTCCACTAGG 22  
 ||||| | |||| | ||||  
 DB 39 AACATGCGGCTCCACTAGG 19

## RESULT 9

RG233190

LOCUS 602667846F1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:482311 5'

DEFINITION mRNA sequence.

ACCESSION RG233190

VERSION RG233190

KEYWORDS EST

SOURCE Homo sapiens (human)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 744)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished

COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgaps@mail.nih.gov](mailto:cgaps@mail.nih.gov)  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (MGRS), Shiroki Toshiyuki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)  
 cDNA sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.llnl.gov>  
 Plate: LLM10732 row: 1 column: 22  
 High quality sequence stop: 741.  
 Location/Qualifiers  
 1..744

## FEATURES

source

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:482313"
/lab_host="DH123"
/clone_lib="NIH_MGC_97"
/notes="Grain: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (Gtcgag
1); Oligo-dT primed using primer 5'-TCTTTTCTTTTCTT-3',
size selected for average insert size 2.2 kb, and
normalized to RCF 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIMH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC library."
BASE COUNT      188 a  179 c  188 g  187 t
ORIGIN
      88.2%; Score 19.4; DB 10; Length 744;
      95.2%; Pred. No. 2.7e-02;
      20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  2  ACCATGGCGGCTCCACTAGG 22
      1  ||||| ||||| ||||| |||||
      12  AACATGGCGGCTCCACTAGG 32

RESULT 12
BF440088      788 bp      mRNA      linear      EST 30 NOV-2000
LOCUS      BF440088
DEFINITION      Maf14 HUVEC cDNA Library Homo sapiens cDNA, mRNA sequence.
ACCESSION      BF440088
VERSION      BF440088.1  GI:11467969
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
    1. bases 1 to 788;
  TITLE      Identification of Genes in HUVES
  JOURNAL      Unpublished
  CONTACT      Contact: Mai H. Nguyen, Canhui Liu
  ORIGIN      Dr. Mai Nguyen
    University of California Los Angeles
    54157 CHS, Surgery Oncology, UCLA, 10833 LeConte Ave, Los Angeles,
    CA 90095, USA
    Tel: 310 267 1923
    Fax: 310 825 7575
    E-mail: mai.nguyen@mednet.ucla.edu; lynch@ucla.mednetnet
    PCR PRIMER 1
    FORWARD  73
    BACKWARD 73
    Insert Length: 788 Std Error: 0.10
    Seq Primer: 73 and 73
    POLY(A)=Yes.

FEATURES
      source          location/qualifiers
      1..788
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /cell_line="Human Umbilical Vein Endothelial Cell"
      /clone_lib="HUVES cDNA Library"
      /note="Vector: Lambda UNIZAP Vector; Site 1: EcoRI;
      Site 2: XhoI"
BASE COUNT      216 a  179 c  194 g  199 t
ORIGIN
      88.2%; Score 19.4; DB 10; Length 788;
      95.2%; Pred. No. 2.7e-02;
      20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  2  ACCATGGCGGCTCCACTAGG 22

```

```

      1  ||||| ||||| ||||| |||||
      9  AACATGGCGGCTCCACTAGG 29

RESULT 11
CB992984      807 bp      mRNA      linear      EST 01 MAY-2003
LOCUS      CB992984
DEFINITION      AGENCOURT 13518822 NIH MGC 148 Homo sapiens cDNA clone
IMAGE:36314747 5', mRNA sequence.
ACCESSION      CB992984
VERSION      CB992984.1  GI:130287504
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
    1. bases 1 to 807;
  TITLE      NIH MGC clone: http://mgc.nci.nih.gov/
  JOURNAL      National Institutes of Health, Mammalian Gene Collection (MGC)
  COMMENT      Unpublished
    Contact: Robert Strausberg, Ph.D.
    Email: cga@nci.nih.gov
    Tissue Procurement: Dr. Stefan Hansson
    cDNA Library Preparation: Michael J. Brownstein (NHGRI) with help
    and advice from Piero Carninci (RIKEN)
    cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
    cDNA Sequencing by: Agencourt Bioscience Corporation
    Clone Distribution: MGC clone distribution information can be
    found through the I.M.A.G.E. Consortium/LINL at:
    http://image.llnl.gov
    Plate: NDW355 row: h column: 20
    High quality sequence stop: 578.
    Location/Qualifiers
      1..807
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="IMAGE:36314747"
      /tissue_type="pre eclamptic placenta"
      /lab_host="DH10B Tera"
      /clone_lib="NIH MGC 148"
      /note="Organ: placenta; Vector: pBluescriptR; Site 1:
      att: XhoI; Site 2: BamH; library is oligo-dT primed and
      directionally cloned using primer
      5'-TTTCTTTTCTTTTCTT-3', size selected for average insert
      size 2.3 kb, and normalized to RCF 5. This is a primary
      library enriched for full-length clones and constructed
      using the Cap-trapper method (Carninci, in preparation).
      Library constructed by M. Brownstein (NIH/NHGRI,
      National Institutes of Health). Note: this is a NIH MGC
      library."
BASE COUNT      216 a  192 c  198 g  201 t
ORIGIN
      88.2%; Score 19.4; DB 14; Length 807;
      95.2%; Pred. No. 2.7e-02;
      20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  2  ACCATGGCGGCTCCACTAGG 22
      1  ||||| ||||| ||||| |||||
      30  AACATGGCGGCTCCACTAGG 50

RESULT 12
BG141322
LOCUS      BG141322
DEFINITION      NG5 HUVEC cDNA Library Homo sapiens cDNA, mRNA sequence.
ACCESSION      BG141322
VERSION      BG141322.1  GI:12643198
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

```



```

RESULT 15
BX437419/c
LOCUS      1022 bp      mRNA      linear      EST 15-MAY-2003
DEFINITION 3-PRIME, mRNA sequence.
ACCESSION  BX437419
VERSION     BX437419.1  GI:30777566
KEYWORDS   EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Mammalia; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 1020)
            Li, W.-L., Gruber, C., Jessen, J., and Polayes, D.
            Full-length cDNA libraries and normalization
            Unpublished
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 101 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 2813.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CSCCAP06B1ONP1cluster-2813.f. Contact :
            Feng Liang Email : fliang@life-techn.com URL :
            http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CSCCAP06B1ONP1.
FEATURES             Location/Qualifiers
     source            1..1020
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /clone="CSCCAP06B1ONP1"
                        /tissue_type="THYMUS"
                        /clone_lib="Homo sapiens THYMUS"
                        /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
                        with a NotI-oligo(dT) primer. Five prime end enriched,
                        double strand cDNA was digested with NotI and cloned into
                        the NotI and EcoRV sites of the pCMVSPORT 6 vector.
                        Library was not normalized."
BASE COUNT      266 a      242 c      242 g      242 t      26 others
ORIGIN
Query Match      88.2%   Score 19.47   DB 13;   Length 1020;
Best local Similarity 94.2%   Pred. No. 2.9e-02;
Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY      2  ACCATGCGCGCTGACCTAGG 25
Dh      243 AACATGCGCGCTGACCTAGG 81

```

Search completed: October 21, 2003, 12:09:48  
 Job time : 124.791 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CorpuGen Ltd.

CM nucleic - nucleic search, using sw model

Run on: October 20, 2003, 21:46:54 : Search time 2.41415 Seconds  
(without alignments)  
3293.972 Million cell updates/sec

Titles: US-10-029 137-9  
Perfect score: 18  
Sequences: 1 tcacgttggttcagag 18

Scoring table: IDENTITY\_NU  
Gapop 10.0, Gapext 1.0

Searched: 56978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 119956

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA :  
1: /cgn2\_6/protdata/2/ina/54\_COMB.seq:  
2: /cgn2\_6/protdata/2/ina/53\_COMB.seq:  
3: /cgn2\_6/protdata/2/ina/52\_COMB.seq:  
4: /cgn2\_6/protdata/2/ina/51\_COMB.seq:  
5: /cgn2\_6/protdata/2/ina/PCMB\_COMB.seq:  
6: /cgn2\_6/protdata/2/ina/backfile1.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17.54	4	540	1	Sequence 1, Appl
2	14.6	82.2	1060	4	Sequence 85, App
3	14.6	82.2	2885	4	Sequence 2, Appl
4	14.6	82.2	2885	4	Sequence 2, Appl
5	14	77.8	2885	4	Sequence 51, Appl
6	14	77.8	2769	4	Sequence 37, Appl
7	14	77.8	2865	4	Sequence 21, Appl
8	14	77.8	2949	4	Sequence 5, Appl
9	14	77.8	3591	4	Sequence 290, App
10	14	77.8	6474	4	Sequence 51, Appl
11	14	77.8	6558	4	Sequence 35, Appl
12	14	77.8	6554	4	Sequence 15, Appl
13	14	77.8	6738	4	Sequence 3, Appl
14	14	77.8	6885	4	Sequence 49, Appl
15	14	77.8	6969	4	Sequence 33, Appl
16	14	77.8	7065	4	Sequence 17, Appl
17	14	77.8	7143	4	Sequence 1, Appl
18	13.8	76.7	97	3	Sequence 15, Appl
19	13.8	76.7	97	3	Sequence 15, Appl
20	13.8	76.7	735	4	Sequence 7959, Ap
21	13.8	76.7	1515	3	Sequence 5, Appl
22	13.8	76.7	1515	3	Sequence 5, Appl
23	13.8	76.7	1660	1	Sequence 3, Appl
24	13.8	76.7	1660	1	Sequence 3, Appl
25	13.8	76.7	1660	5	Sequence 11, Appl
26	13.8	76.7	1750	2	Sequence 11, Appl
27	13.8	76.7	1750	2	Sequence 11, Appl

28	13.8	76.7	2312	1	US-08-132-942A-1	Sequence 1, Appl
29	13.8	76.7	2312	4	US-09-037-179B-1	Sequence 1, Appl
30	13.8	76.7	3020	4	US-09-220-132-19	Sequence 19, App
31	13.8	76.7	3358	3	US-09-248-571-2	Sequence 2, Appl
32	13.8	76.7	3358	4	US-09-553-736-2	Sequence 2, Appl
33	13.8	76.7	3552	4	US-09-107-532A-3026	Sequence 3026, Ap
34	13.8	76.7	6749	4	US-08-961-527-84	Sequence 84, Appl
35	13.8	76.7	8460	1	US-08-469-055A-9	Sequence 9, Appl
36	13.8	76.7	8519	3	US-09-261-927-1	Sequence 1, Appl
37	13.8	76.7	42157	4	US-08-311-731A-126	Sequence 126, App
38	13.8	76.7	42572	4	US-08-340-347-3	Sequence 3, Appl
39	13.4	74.4	120	1	US-08-189-236-1	Sequence 1, Appl
40	13.4	74.4	834	3	US-09-176-657-6	Sequence 6, Appl
41	13.4	74.4	834	4	US-09-421-299-6	Sequence 6, Appl
42	13.4	74.4	1350	1	US-07-792-466-7	Sequence 7, Appl
43	13.4	74.4	1350	1	US-08-424-633A-1	Sequence 1, Appl
44	13.4	74.4	1350	2	US-08-737-524B-4	Sequence 4, Appl
45	13.4	74.4	1350	4	US-08-723-771-1	Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-09-309-317 1/c  
Sequence 1, Application US/09309317  
Patent No. 6277970  
GENERAL INFORMATION:  
APPLICANT: Prusiner, Stanley  
APPLICANT: Tremblay, Patrick  
APPLICANT: Moore, Richard  
APPLICANT: Westaway, David  
APPLICANT: Hood, Leroy E.  
APPLICANT: Lee, Inyong  
TITLE OF INVENTION: Prp-like Gene  
FILE REFERENCE: 6510-130US1  
CURRENT APPLICATION NUMBER: US/09/309,317  
CURRENT FILING DATE: 1999-05-11  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 540  
TYPE: CNA  
ORGANISM: mus musculus  
US 09-309-317-1

Query Match: 94.4% Score 17 DB ID: Length 540  
Best Local Similarity: 100.0%, Pred. No. 218  
Matches: 17 Conservative 0 Mismatches 0 Gaps 0

QY : TCACGTGCTTCAGAG 17  
DB 301 TCACGTGCTTCAGAG 285

RESULT 2  
US-09-620-3129-853/C  
Sequence 853, Application US/096203129  
Patent No. 6569662  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhang, Jie  
APPLICANT: Ren, Felyan  
APPLICANT: Chen, Rui-hong  
APPLICANT: Zhao, Qing A.  
APPLICANT: Weinman, Tom  
APPLICANT: Xue, Aigang J.  
APPLICANT: Yang, Yonghong  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Zhou, Ping  
APPLICANT: Ma, Yunjing

```

; APPLICANT: Wang, Dunru
; APPLICANT: Wang, Zhwei
; APPLICANT: Cao, Tillinghast
; APPLICANT: Ormanac, Radote T.
; TITLE OF INVENTION: POLYPEPTIDES
; FILE REFERENCE: 784C1F2B
; CURRENT APPLICATION NUMBER: US/09/623,312D
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1206
; SOFTWARE: p1.1.1 Genes Version 1.0
; SEQ ID NO 63
; LENGTH: 1567
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURES:
; NAME/KEY: CDS
; LOCATION: 1151..1984;
US-09-623-312D-633

```

```

Query Match      82.2%; Score 14.8; DB 3; Length 1567;
Best Local Similarity 88.9%; Pred. No. 63;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 TCACCTTGGCTTCACAGG 18
DB 218 TCACCTTGGCTTCACAGG 711

```

```

RESULT 3
US-08-948-705-270
; Sequence 2, Application US/08948705A
; Patent No. 6041054
; GENERAL INFORMATION:
; APPLICANT: Scallan, Matthew J.
; APPLICANT: Chen, Yao-Tseeg
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ASSOCIATED WITH COLON CANCER AND METHODS FOR DIAGNOSING AND
; FILE REFERENCE: LJD-5506-JEL/NDH
; CURRENT APPLICATION NUMBER: US/08/948,705A
; CURRENT FILING DATE: 1997-10-10
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2885
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-08-948-705-2

```

```

Query Match      52.2%; Score 14.8; DB 3; Length 2885;
Best Local Similarity 62.9%; Pred. No. 63;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 TCACCTTGGCTTCACAGG 18
DB 227 TCACCTTGGCTTCACAGG 210

```

```

RESULT 4
US-09-510-543-270
; Sequence 2, Application US/09510543
; Patent No. 6517837
; GENERAL INFORMATION:
; APPLICANT: Scallan, Matthew J.
; APPLICANT: Chen, Yao-Tseeg
; APPLICANT: Stockert, Elisabeth

```

```

; APPLICANT: Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ASSOCIATED WITH COLON CANCER
; FILE REFERENCE: LJD-5506-JEL/NDH
; CURRENT APPLICATION NUMBER: US/09/510,543
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 09/120,120
; PRIOR FILING DATE: 1999-06-22
; NUMBER OF SEQ ID NOS: 22
; SEQ ID NO 2
; LENGTH: 2885
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-510-543-2

```

```

Query Match      77.6%; Score 14.7; DB 4; Length 2885;
Best Local Similarity 106.3%; Pred. No. 170+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 TCACCTTGGCTTCACAGG 18
DB 227 TCACCTTGGCTTCACAGG 210

```

```

RESULT 5
US-09-854-856-53
; Sequence 53, Application US/09854856
; Patent No. 6541152
; GENERAL INFORMATION:
; APPLICANT: Walker, D. Wade
; APPLICANT: Hilburn, Rich
; APPLICANT: Decker, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: Human Kinases and Polynucleotides
; FILE REFERENCE: LEX-0178 USA
; CURRENT APPLICATION NUMBER: US/09/854,856
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/206,015
; PRIOR FILING DATE: 2001-05-19
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 53
; LENGTH: 2685
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-854-856-53

```

```

Query Match      77.6%; Score 14.7; DB 4; Length 2685;
Best Local Similarity 106.3%; Pred. No. 170+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 TCACCTTGGCTTCACAGG 18
DB 257 TCACCTTGGCTTCACAGG 210

```

```

RESULT 6
US-09-854-856-37
; Sequence 37, Application US/09854856
; Patent No. 6541152
; GENERAL INFORMATION:
; APPLICANT: Walker, D. Wade
; APPLICANT: Hilburn, Rich
; APPLICANT: Decker, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: Human Kinases and Polynucleotides
; FILE REFERENCE: LEX-0178 USA
; CURRENT APPLICATION NUMBER: US/09/854,856
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/206,015
; PRIOR FILING DATE: 2001-05-19

```

1 NUMBER OF SEQ ID NOS: 64  
2 SOFTWARE: FastSeq for Windows Version 4.0  
3 SEQ ID NO 17  
4 LENGTH: 2949  
5 TYPE: DNA  
6 ORGANISM: Homo sapiens  
US 09-854-856-17

Query Match 77.8% Score 14; DB 4; Length 2769;  
Best Local Similarity 100.0%; Pred. No. 1.7e-02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAGTTGGCTTCA 14  
||| ||| ||| |||  
DB 2411 TCAGTTGGCTTCA 2464

RESULT 9  
US 09-854-856-21  
1 Sequence 290, Application US/09854856  
2 Patent No. 6541252  
3 GENERAL INFORMATION:  
4 APPLICANT: Walker, D. Wade  
5 APPLICANT: Hilborn, Eric  
6 APPLICANT: Donoho, Gregory  
7 TITLE OF INVENTION: Turner, C. Alexander, Jr.  
8 TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides  
9 FILE REFERENCE: LEX-0178 USA  
10 CURRENT APPLICATION NUMBER: US/09/854-856  
11 PRIOR FILING DATE: 2001-05-14  
12 PRIOR APPLICATION NUMBER: US 60/206,015  
13 PRIOR FILING DATE: 2000-05-19  
14 NUMBER OF SEQ ID NOS: 64  
15 SOFTWARE: FastSeq for Windows Version 4.0  
16 SEQ ID NO 21  
17 LENGTH: 2865  
18 TYPE: RNA  
19 ORGANISM: Homo sapiens  
US 09-854-856-21

Query Match 77.8% Score 14; DB 4; Length 2865;  
Best Local Similarity 100.0%; Pred. No. 1.7e-02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAGTTGGCTTCA 14  
||| ||| ||| |||  
DB 2517 TCAGTTGGCTTCA 2350

RESULT 8  
US 09-854-856-5  
1 Sequence 57, Application US/09854856  
2 Patent No. 6541252  
3 GENERAL INFORMATION:  
4 APPLICANT: Walker, D. Wade  
5 APPLICANT: Hilborn, Eric  
6 APPLICANT: Donoho, Gregory  
7 TITLE OF INVENTION: Turner, C. Alexander, Jr.  
8 TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides  
9 FILE REFERENCE: LEX-0178 USA  
10 CURRENT APPLICATION NUMBER: US/09/854-856  
11 PRIOR FILING DATE: 2001-05-14  
12 PRIOR APPLICATION NUMBER: US 60/206,015  
13 PRIOR FILING DATE: 2000-05-19  
14 NUMBER OF SEQ ID NOS: 64  
15 SOFTWARE: FastSeq for Windows Version 4.0  
16 SEQ ID NO 5  
17 LENGTH: 2949  
18 TYPE: RNA  
19 ORGANISM: Homo sapiens  
US 09-854-856-5

Query Match 77.8% Score 14; DB 4; Length 2949;  
Best Local Similarity 100.0%; Pred. No. 1.7e-02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAGTTGGCTTCA 14  
||| ||| ||| |||  
DB 2621 TCAGTTGGCTTCA 2634

RESULT 9  
US 09-107-512A-290-0  
1 Sequence 290, Application US/09107512A  
2 Patent No. 6583275  
3 GENERAL INFORMATION:  
4 APPLICANT: Lyryo A. Daccette Starr and David Bush  
5 TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
6 NUMBER OF SEQUENCES: 7310  
7 CORRESPONDENCE ADDRESS:  
8 ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
9 STREET: 150 Beaver Street  
10 CITY: Waltham  
11 STATE: Massachusetts  
12 COUNTRY: USA  
13 ZIP: 02354  
14 COMPUTER READABLE FORM:  
15 MEDIUM TYPE: CD-ROM ISO9660  
16 COMPUTER: PC  
17 OPERATING SYSTEM: <UNKNOWN>  
18 SOFTWARE: ASCII  
19 CURRENT APPLICATION DATA:  
20 APPLICATION NUMBER: US/09/107,512A  
21 FILING DATE: 30 Jun 1998  
22 PRIOR APPLICATION DATA:  
23 APPLICATION NUMBER: 60/085,598  
24 FILING DATE: 14 May 1998  
25 APPLICATION NUMBER: 60/085,571  
26 FILING DATE: July 2, 1997  
27 ATTORNEY/AGENT INFORMATION:  
28 NAME: Arieliello, Pamela Deneke  
29 REGISTRATION NUMBER: 40,499  
30 REFERENCE/DOCKET NUMBER: STC-012  
31 TELECOMMUNICATION INFORMATION:  
32 TELEPHONE: (781)893-5007  
33 TELEFAX: (781)893-9277  
34 INFORMATION FOR SEQ ID NO: 290:  
35 SEQUENCE CHARACTERISTICS:  
36 LENGTH: 3591 base pairs  
37 TYPE: nucleic acid  
38 STRANDEDNESS: double  
39 TOPOLOGY: circular  
40 MOLECULE TYPE: DNA (genomic)  
41 HYPOTHETICAL: NO  
42 ANTI-SENSE: NO  
43 ORIGINAL SOURCE:  
44 ORGANISM: Enterococcus faecium  
45 FEATURE:  
46 NAME/KEY: misc feature  
47 LOCATION: (1) LOCATION 1...3591;  
48 SEQUENCE DESCRIPTION: SEQ ID NO: 290:  
US-09-107-512A-290

Query Match 77.8% Score 14; DB 4; Length 3591;  
Best Local Similarity 100.0%; Pred. No. 1.8e-02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAGTTGGCTTCA 14  
||| ||| ||| |||  
DB 3397 TCAGTTGGCTTCA 3184

RESULT 10

Tu Oct 21 11:34:44 2003

```

US-09-854-856-51
; Sequence 51, Application US/C9854856
; Patent No. 6541252
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Hilburn, Erin
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
; FILE REFERENCE: LEX-0178-USA
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/206,015
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 51
; LENGTH: 6474
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-854-856-51

Query Match 77.8%; Score 14; DB 4; Length 6474;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1 TCACGTTGGCTTCA 14
DB 2357 TCACGTTGGCTTCA 2370

RESULT 11
US-09-854-856-35
; Sequence 35, Application US/09854856
; Patent No. 6541252
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Hilburn, Erin
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
; FILE REFERENCE: LEX-0178-USA
; CURRENT APPLICATION NUMBER: US/09/854,856
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/206,015
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 6558
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-854-856-35

Query Match 77.8%; Score 14; DB 4; Length 6558;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1 TCACGTTGGCTTCA 14
DB 2441 TCACGTTGGCTTCA 2454

RESULT 12
US-09-854-856-19
; Sequence 19, Application US/09854856
; Patent No. 6541252
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Hilburn, Erin
; APPLICANT: Donoho, Gregory

```

```

; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
; FILE REFERENCE: LEX-0178-USA
; CURRENT APPLICATION NUMBER: US/09/854,856
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/206,015
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 6554
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-854-856-19

Query Match 77.8%; Score 14; DB 4; Length 6554;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1 TCACGTTGGCTTCA 14
DB 2537 TCACGTTGGCTTCA 2550

RESULT 13
US-09-854-856-3
; Sequence 3, Application US/09854856
; Patent No. 6541252
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Hilburn, Erin
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
; FILE REFERENCE: LEX-0178-USA
; CURRENT APPLICATION NUMBER: US/09/854,856
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/206,015
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 6738
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-854-856-3

Query Match 77.8%; Score 14; DB 4; Length 6738;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1 TCACGTTGGCTTCA 14
DB 2621 TCACGTTGGCTTCA 2634

RESULT 14
US-09-854-856-49
; Sequence 49, Application US/09854856
; Patent No. 6541252
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Hilburn, Erin
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
; FILE REFERENCE: LEX-0178-USA
; CURRENT APPLICATION NUMBER: US/09/854,856
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/206,015

```



? PRIOR FILING DATE: 2000-05-19  
 ? NUMBER OF SEQ ID NOS: 64  
 ? SOFTWARE: FastSeq for Windows Version 4.0  
 ? SEQ ID NO 49  
 ? LENGTH: 6885  
 ? TYPE: DNA  
 ? ORGANISM: Homo sapiens  
 US-09-854-856-49

Query Match 71.8% Score 14; DB 4; Length 6885;  
 Best Local Similarity 100.0%; Pred. No. 2e-02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACGTTGGCTTCA 14  
 DE 2457 TCACGTTGGCTTCA 2457

## RESULT 15

US-09-854-856-13  
 ? Sequence 13, Application US/9854856  
 ? Patent No. 6541252  
 ? GENERAL INFORMATION:  
 ? APPLICANT: Walke, D. Wade  
 ? APPLICANT: Hilborn, Eric  
 ? APPLICANT: Donoso, Gregory  
 ? APPLICANT: Turner, C. Alexander Jr.  
 ? TITLE OF INVENTION: NO. 6541252: Human Kinases and Polynucleotides  
 ? FILE REFERENCE: EX-0178-USA  
 ? CURRENT APPLICATION NUMBER: US/99/854-856  
 ? CURRENT FILING DATE: 2001-05-19  
 ? PRIOR APPLICATION NUMBER: US 60/206,016  
 ? PRIOR FILING DATE: 2000-05-19  
 ? NUMBER OF SEQ ID NOS: 64  
 ? SOFTWARE: FastSeq for Windows Version 4.0  
 ? SEQ ID NO 33  
 ? LENGTH: 6869  
 ? TYPE: DNA  
 ? ORGANISM: Homo sapiens  
 US-09-854-856-13

Query Match 71.8% Score 14; DB 4; Length 6869;  
 Best Local Similarity 100.0%; Pred. No. 2e-02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACGTTGGCTTCA 14  
 DE 2441 TCACGTTGGCTTCA 2454

Search completed: October 21, 2003, 11:34:44  
 Job file: 244115.ncs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

GV nucleic - nucleic search, using sw model

Run on: October 20, 2003, 11:12:34 : Search time 114.401 seconds  
without alignment:  
7537.992 Million cell updates/sec

Title: US-10-029-137-10  
Perfect scores: 22  
Sequences: 1: caaccatgagctccacaaaga 22

Scoring table: IDENTITY.MD

Gapop 10.0, Gapext 1.0

Searches: 2888/11 seqs, 204649186 residues

Total number of hits satisfying chosen parameters: 577423

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum March 03  
Maximum March 1003  
Display first 45 summaries

Database: GenBank

1: gb\_bat

2: gb\_htg

3: gb\_hni

4: gb\_om

5: gb\_ov

6: gb\_pat

7: gb\_ph

8: gb\_pri

9: gb\_pst

10: gb\_ric

11: gb\_sis

12: gb\_sy

13: gb\_tus

14: gb\_vit

15: gb\_ba

16: em\_fun

17: em\_hu

18: em\_in

19: em\_mus

20: em\_mus

21: em\_mus

22: em\_ov

23: em\_pat

24: em\_ph

25: em\_pst

26: em\_ric

27: em\_sis

28: em\_tus

29: em\_vit

30: em\_htg\_hu

31: em\_htg\_invi

32: em\_htg\_other

33: em\_htg\_mus

34: em\_htg\_pri

35: em\_htg\_ric

36: em\_htg\_mam

37: em\_htg\_vit

38: em\_sy

39: em\_htg\_hu

40: em\_htg\_mus

41: em\_htg\_other

Prod. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	EB	ID	Description
1	13.4	88.2	455	1	BD025539	BD025539 Sequence
2	13.4	88.2	457	1	BD025539	BD025539 EST and e
3	13.4	88.2	783	3	AF176800	AF176800 Homo sapi
4	19.4	86.2	522	6	BD015361	BD015361 Human pro
5	19.4	88.2	882	6	AX014817	AX014817 Sequence
6	19.4	88.2	1071	6	AX017517	AX017517 Sequence
7	19.4	88.2	1071	6	BC135207	BC135207 Human bac
8	19.4	88.2	1083	9	BC011936	BC011936 Homo sapi
9	19.4	88.2	1083	6	AX015550	AX015550 Sequence
10	19.4	88.2	1353	9	AF158829	AF158829 Homo sapi
11	19.4	88.2	3109	9	AF121617	AF121617 Homo sapi
12	19.4	88.2	179310	9	AC006160	AC006160 Homo sapi
13	18.9	85.5	209293	2	AC111176	AC111176 Rattus no
14	18	81.8	177293	2	AC125455	AC125455 Mus muscu
15	18	81.8	126923	10	AC108948	AC108948 Mus muscu
16	17.8	81.8	201533	10	AC124483	AC124483 Mouse DNA
17	17.8	80.9	68954	9	AC125532	AC125532 Homo sapi
18	17.8	80.9	93255	10	AF002736	AF002736 Mus muscu
19	17.8	80.9	182198	2	BX121894	BX121894 Dario rei
20	17.8	80.9	181131	2	AC101865	AC101865 Mus muscu
21	17.8	80.9	196472	7	FX055377	FX055377 Dario rei
22	17.8	80.9	198593	10	AC015800	AC015800 Mus muscu
23	17.8	80.9	198739	2	BX005307	BX005307 Dario rei
24	17.8	80.9	212766	2	AC121130	AC121130 Rattus no
25	17.8	80.9	223118	2	AC094074	AC094074 Rattus no
26	17.8	80.9	223873	9	AC008735	AC008735 Homo sapi
27	17.8	80.9	249487	10	VX0251768	VX0251768 Mus muscu
28	17.8	80.9	249357	10	VX0251835	VX0251835 Mus muscu
29	17.8	80.9	275713	2	AC097881	AC097881 Rattus no
30	17.4	79.1	244	6	BD016709	BD016709 Silkwo
31	17.2	78.2	1519	4	BT022309	BT022309 Arabidops
32	17.2	78.2	1899	1	AY142767	AY142767 Heliohaci
33	17.2	78.2	2314	3	AY133810	AY133810 Arabidops
34	17.2	78.2	2452	4	AY020017	AY020017 Arabidops
35	17.2	78.2	60164	2	AC020373	AC020373 Drosophi
36	17.2	78.2	66348	2	AC100317	AC100317 Homo sapi
37	17.2	78.2	85518	1	AC008549	AC008549 Drosophi
38	17.2	78.2	100665	8	AC006533	AC006533 Arabidops
39	17.2	78.2	145020	2	AC114605	AC114605 Mus muscu
40	17.2	78.2	145155	9	AC091233	AC091233 Oryza sat
41	17.2	78.2	161500	3	AC103736	AC103736 Homo sapi
42	17.2	78.2	17576	10	AC120049	AC120049 Mus muscu
43	17.2	78.2	172843	10	AC134452	AC134452 Mouse DNA
44	17.2	78.2	194034	7	AC091348	AC091348 Drosophi
45	17.2	78.2	194034	7	AC091348	AC091348 Drosophi

ALIGNMENTS

RESULT :	BD025539	Sequence tag and encoded human protein.	465 bp	DNA	Linear	PAT 27 AUG-2002
LOCUS	BD025539					
DEFINITION	BD025539					
ACCESSION	BD025539.1	GI:22556762				
VERSION	CP 2001269.82-A/1785					
KEYWORDS	Homo sapiens (human)					
SOURCE	Homo sapiens					
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE	1. (bases 1 to 465)					
AUTHORS	Edwards, J.B.D.M., Chelazzi, E. and Jordan, J.Y.					
DATE	Sequence tag and encoded human protein					
JOURNAL	Patent: CP 0001269.82-A 1785 02-OCT-2001					

GENSET:  
OS Homo sapiens (human);  
PN JP 2001269182-A/1785  
PD 22-OCT-2001  
PF 24-FEB-2000 JP 2000118773  
PR 26-FEB-1999 US 60/122487  
PI JEAN BAPTISTE DUMAS MICHE EDWARDS, EMMERIC DUCLAIR, JEAN YVES  
PC C12N15/09, C07K14/435, C07K16/18, C12N1/15, C12N1/19, C12N1/21, PC  
C12N5/10, C12N1/22, C12P21/02, C12Q1/28, C12Q1/68, C12N15/00, C12N5/00, PC  
G06F15/40

CC  
FH Key Location/Qualifiers  
FT CDS 20..463.  
Location/Qualifiers  
1..465  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

BASE COUNT 116 a 113 c 122 g 128 t 5 others  
ORIGIN

Query Match 88.2%; Score 19.4; DB 6; Length 465;  
Best Local Similarity 95.2%; Pred. No. 45;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AACATGGCGGCTCCACTAGG 22  
DB 17 AACATGGCGGCTCCACTAGG 37

RESULT 2  
BD108605  
LOCUS EST and encoded human protein.  
DEFINITION BD108605  
ACCESSION BD108605.1 GI:23203423  
VERSION JP 2002010789-A/692  
KEYWORDS Homo sapiens (human);  
SOURCE Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
Edwards, J.B.D.M., Robert, S. and Giordano, J.E.  
EST and encoded human protein  
Patent: JP 2002010789-A 692 15 JAN-2002;  
GENSET CORP

COMMENT  
OS Homo sapiens (human);  
PN JP 2002010789-A/692  
PD 15 JAN-2002  
PF 07-AUG-2000 JP 2000280989  
PR 05-AUG-1999 US 60/124749  
PI JEAN BAPTISTE DUMAS MICHE EDWARDS, SEVEILIN JOBERT, JEAN YVES  
GIORDANO  
PC C12N15/09, C12N1/21, C07K14/47, C07K16/18, C12N1/15, C12N1/19, PC  
C12N5/10, C12P21/02, C12P21/08, C12Q1/68, C12N15/00, C12N5/00, PC  
C12N15/00  
CC EST and encoded human protein  
FH Key Location/Qualifiers  
FT CDS 15..467.  
Location/Qualifiers  
1..467  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

BASE COUNT 121 a 113 c 123 g 107 t 3 others  
ORIGIN

Query Match 88.2%; Score 19.4; DB 6; Length 467;  
Best Local Similarity 95.2%; Pred. No. 45;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AACATGGCGGCTCCACTAGG 22  
DB 12 AACATGGCGGCTCCACTAGG 37

RESULT 3  
BD108605  
LOCUS EST and encoded human protein.  
DEFINITION BD108605  
ACCESSION BD108605.1 GI:23203423  
VERSION JP 2002010789-A/692  
KEYWORDS Homo sapiens (human);  
SOURCE Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
Edwards, J.B.D.M., Robert, S. and Giordano, J.E.  
EST and encoded human protein  
Patent: JP 2002010789-A 692 15 JAN-2002;  
GENSET CORP

COMMENT  
OS Homo sapiens (human);  
PN JP 2002010789-A/692  
PD 15 JAN-2002  
PF 07-AUG-2000 JP 2000280989  
PR 05-AUG-1999 US 60/124749  
PI JEAN BAPTISTE DUMAS MICHE EDWARDS, SEVEILIN JOBERT, JEAN YVES  
GIORDANO  
PC C12N15/09, C12N1/21, C07K14/47, C07K16/18, C12N1/15, C12N1/19, PC  
C12N5/10, C12P21/02, C12P21/08, C12Q1/68, C12N15/00, C12N5/00, PC  
C12N15/00  
CC EST and encoded human protein  
FH Key Location/Qualifiers  
FT CDS 15..467.  
Location/Qualifiers  
1..467  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

BASE COUNT 121 a 113 c 123 g 107 t 3 others  
ORIGIN

Query Match 88.2%; Score 19.4; DB 6; Length 467;  
Best Local Similarity 95.2%; Pred. No. 42;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AACATGGCGGCTCCACTAGG 22  
DB 9 AACATGGCGGCTCCACTAGG 29

RESULT 4  
BD013361  
LOCUS Human protein and cDNA.  
DEFINITION BD013361  
ACCESSION BD013361.1 GI:22556409  
VERSION JP 2001218584-A/7  
KEYWORDS Homo sapiens (human);  
SOURCE Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
Kato, S. and Sackl, N.  
Human protein and cDNA(4)  
Patent: JP 2001218584-A 7 14 AUG-2001;  
JAPAN SCIENCE AND TECHNOLOGY CORP  
OS Homo sapiens (human);  
PN JP 2001218584-A/7  
PD 14-AUG-2001  
PF 08-FEB-2000 JP 2000031062

```

P1 SETCI KATO,MIROPO SAEKI
PC C12N15/09,C07K14/435,C07K16/18,C12N1/15,C12N1/19,C12N/21, PC
C12N5/10/7
PC C12P21/08,C12N15/05,C12N5/00
CC
PI Key Location/Qualifiers
PT CDS Location/Qualifiers
FEATURES
    source
        1..922
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
BASE COUNT      210 a 186 c 193 g 127 t
ORIGIN
    Query Match      64.2%; Score 19.4; DB 6; Length 922.
    Best Local Similarity 95.2%; Pred. No. 42;
    Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

CY 2 ACCATGGCGGCTCCACTAGG 22
   ||||| ||||| |||||
DB 7 AACATGGCGGCTCCACTAGG 27

RESULT 5
LOCUS      AX014817      883 bp      DNA      linear      PAT 07-SEP-2002
DEFINITION Sequence 3 from Patent WO9503040.
ACCESSION  AX014817
VERSION     AX014817.1  GI:10041284
KEYWORDS   Homo sapiens (human)
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  1. 882
AUTHORS    Schmidt,A., Specht,T., Duhl,E., Hinzmman,B., Rosenthal,A. and
Pilarczyk,C.
TITLE      Human nucleic acid sequences from ovarian tumour tissue
JOURNAL    Patent: WO 9503040 A 3 21 OCT 1993;
            SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
            BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
            (DE); PILARSKY CHRISTIAN (DE)
FEATURES   Location/Qualifiers
            1..882
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
BASE COUNT      210 a 186 c 193 g 127 t
ORIGIN
    Query Match      64.2%; Score 19.4; DB 6; Length 922.
    Best Local Similarity 95.2%; Pred. No. 42;
    Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

CY 2 ACCATGGCGGCTCCACTAGG 22
   ||||| ||||| |||||
DB 7 AACATGGCGGCTCCACTAGG 27

RESULT 6
LOCUS      AX017517      1071 bp      DNA      linear      PAT 07-SEP-2002
DEFINITION Sequence 65 from Patent WO947455.
ACCESSION  AX017517
VERSION     AX017517.1  GI:10042314
KEYWORDS   Homo sapiens (human)
SOURCE     Homo sapiens
ORGANISM   Homo sapiens
REFERENCE  1
AUTHORS    Schmidt,A., Specht,T., Dahl,E., Hinzmman,B., Rosenthal,A. and
Pilarczyk,C.
TITLE      Human nucleic acid sequences from normal breast tissue
JOURNAL    Patent: WO 947455 A 65 23 SEP 1999;
            SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
            BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
            (DE); PILARSKY CHRISTIAN (DE)
FEATURES   Location/Qualifiers
            1..1071
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
BASE COUNT      281 a 218 c 255 g 317 t
ORIGIN
    Query Match      88.2%; Score 19.4; DB 6; Length 1071;
    Best Local Similarity 95.2%; Pred. No. 40;
    Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

CY 2 ACCATGGCGGCTCCACTAGG 22
   ||||| ||||| |||||
DB 7 AACATGGCGGCTCCACTAGG 27

```

```

Pilarczyk,C.
Human nucleic acid sequences from normal breast tissue
Patent: WO 947455 A 65 23 SEP 1999;
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILARSKY CHRISTIAN (DE)
FEATURES   Location/Qualifiers
            1..1071
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
BASE COUNT      281 a 218 c 255 g 317 t
ORIGIN
    Query Match      88.2%; Score 19.4; DB 6; Length 1071;
    Best Local Similarity 95.2%; Pred. No. 40;
    Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

CY 2 ACCATGGCGGCTCCACTAGG 22
   ||||| ||||| |||||
DB 7 AACATGGCGGCTCCACTAGG 27

RESULT 7
LOCUS      BD135207      1071 bp      DNA      linear      PAT 18-SEP-2002
DEFINITION Human nucleic acid sequence originating in normal mammary tissue.
ACCESSION  BD135207
VERSION     BD135207.1  GI:23230152
KEYWORDS   Homo sapiens (human)
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  1. 1071
AUTHORS    Specht,T., Hinzmman,B., Schmitt,A., Pirarski,C., Duhl,E. and
Rosenthal,A.
TITLE      Human nucleic acid sequence originating in normal mammary tissue
JOURNAL    Patent: JP 2002506639-A 54 05 MAR 2002;
            METAGEN GESELLSCHAFT FUER GENOM FÖRSCHUNG MBH
            OS Homo sapiens (human)
COMMENT     PN JP 2002506639-A/54
            PC 05-MAR-2002
            PF 19-MAR-1999 JP 2000536838
            PR 20-MAR-1998 DE 198 13 835.0
            PI THOMAS SPECHT,BERND HINZMAN,ARMIN SCHMITT,CHRISTIAN PIRARSKI,
            EDGAR DÜHL,
            PI ANDRE ROSENTHAL
            PC C12N15/09,A61K48/00,A61P35/00,A61P43/00,A61P43/50,C07K14/47,
            C07K16/18,
            PC C12N1/21,C12N5/10,C12P21/02,C12Q1/68,G01K33/68//A61K38/00, PC
            C12N5/00,A61K47/62
            PC C12N5/00,A61K47/62
            CC Human nucleic acid sequence originating in normal mammary CC
            tissue
FEATURES   Location/Qualifiers
            1..1071
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
BASE COUNT      281 a 218 c 255 g 317 t
ORIGIN
    Query Match      88.2%; Score 19.4; DB 6; Length 1071;
    Best Local Similarity 95.2%; Pred. No. 40;
    Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

CY 2 ACCATGGCGGCTCCACTAGG 22
   ||||| ||||| |||||
DB 7 AACATGGCGGCTCCACTAGG 27

```

RESULT 8  
BC011936  
LOCUS  
DEFINITION Homo sapiens, hypothetical, protein DKFZp434N195, clone MGC:25562  
IMAGE:3936713, mRNA, complete cds.  
ACCESSION BC011936  
VERSION BC011936.1 GI:15080366  
KEYWORDS MGC  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 1083)  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (30 JUL 2001) National Institutes of Health, National Cancer  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A01, Bethesda, MD 20892-1500,  
USA  
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
Tissue Procurement: DCTD/OTP  
cDNA library Preparation: Rubin Laboratory  
cDNA library Arrayed by: The I.M.A.G.E. Consortium (LMML)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@hgrl.nih.gov](mailto:nisc_mgc@hgrl.nih.gov)  
Shenchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,  
Benjamin, B., Blakesley, R.W., Bouffard, G.S., Brinkley, C., Brooks, S.,  
Dietrich, N.S., Guan, C., Gupta, S., Ho, S.-L., Karlins, E., Legaspi, P.,  
Lim, M., Maduro, Q., Masello, C., Mastrian, S.D., McCloskey, J.C.,  
McDowell, J., Pearson, R., Snyder, B., Stantipop, S., Thomas, P.J.,  
Tiongson, E.E., Touchman, J.W., Tsargov, C., Vogt, J.L., Walker, M.A.,  
Zhang, S.-H., and Green, S.D.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLM at: <http://image.llnl.gov>  
Series: IMAGE Plate: 28 Row: 6 Column: 4  
This clone was selected for full length sequencing because it  
passed the following selection criteria: Hexamer frequency ORF  
analysis.

FEATURES  
source  
1..1083  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="LOCUSID:80106"  
/db\_xref="taxon:9606"  
/clone="MGC:25562 IMAGE:3936713"  
/tissue\_type="lung, small cell carcinoma"  
/clone\_lib="NIH\_MGC\_7"  
/lab\_host="DH10B-R"  
/note="Vector: pOT57"  
24..550  
/codon\_start=1  
/product="hypothetical protein DKFZp434N195"  
/protein\_id="AAH1936.1"  
/db\_xref="GI:15080366"  
/translation="MAAPLGGMFSGQPPQAPPGLPGQASLQAAPGAPRPSSTL  
VDELSSEAFASLVSDVNGTQEEIRTGVDCCIOKFLDIAROTETCFELQKRLQ  
SVQKPEQVIRKDSVSELSNELQRDALVQKHLTKLRHWQVLEINVQHKKPADIPGS  
LAYLEQASANIIPALEKPT"  
BASE COUNT 294 a 216 c 259 g 314 t  
ORIGIN  
Query Match 88.2% Score 19.4; DB 9; Length 1083;  
Best Local Similarity 95.2% Pred. No. 40;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 9  
AX015050  
LOCUS  
DEFINITION Homo sapiens, hypothetical, protein DKFZp434N195, clone MGC:25562  
IMAGE:3936713, mRNA, complete cds.  
ACCESSION AX015050  
VERSION AX015050.1 GI:15080366  
KEYWORDS MGC  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 1083)  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (30 JUL 2001) National Institutes of Health, National Cancer  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A01, Bethesda, MD 20892-1500,  
USA  
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
Tissue Procurement: DCTD/OTP  
cDNA library Preparation: Rubin Laboratory  
cDNA library Arrayed by: The I.M.A.G.E. Consortium (LMML)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@hgrl.nih.gov](mailto:nisc_mgc@hgrl.nih.gov)  
Shenchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,  
Benjamin, B., Blakesley, R.W., Bouffard, G.S., Brinkley, C., Brooks, S.,  
Dietrich, N.S., Guan, C., Gupta, S., Ho, S.-L., Karlins, E., Legaspi, P.,  
Lim, M., Maduro, Q., Masello, C., Mastrian, S.D., McCloskey, J.C.,  
McDowell, J., Pearson, R., Snyder, B., Stantipop, S., Thomas, P.J.,  
Tiongson, E.E., Touchman, J.W., Tsargov, C., Vogt, J.L., Walker, M.A.,  
Zhang, S.-H., and Green, S.D.

RESULT 10  
AF358829  
LOCUS  
DEFINITION Homo sapiens, tumor angiogenesis marker mRNA, complete cds.  
ACCESSION AF358829  
VERSION AF358829.1 GI:13635459  
KEYWORDS MGC  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 1363)  
AUTHORS Liu, C., Zhang, J., Shao, Z.M., Bratty, P., Sathappan, M., Lane, T.F.,  
Baskys, S.H., Livingston, E., and Nguyen, M.  
TITLE Identification of a novel endothelial-derived gene EG-1  
Biochem Biophys Res Commun 296 (1), 612-612 (2002)  
21638093  
MEDLINE  
PUBMED 11774215  
REFERENCE 2 (bases 1 to 1363)  
AUTHORS Nguyen, M.H. and Liu, C.  
TITLE A Novel Gene as Tumor Angiogenesis Marker  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 1363)  
AUTHORS Nguyen, M.H. and Liu, C.  
TITLE Direct Submission  
JOURNAL Submitted (08 MAR 2001) Surgical Oncology, UCLA Medical School,  
10833 Le Conte Ave., Los Angeles, CA 90095, USA  
10833 Le Conte Ave., Los Angeles, CA 90095, USA  
11774215  
FEATURES  
source  
1..1363  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"

```

CDS
13...549
/codon_start=1
/product="tumor angiogenesis marker"
/protein_id="AAK32724.1"
/db_xref="GI:13605460"
/translation="MAAPLGWVFSCQRPQPCAPPGLRQASHLQAAFGAPRPSSSTL
VDELSFACASVSCDYVNTQDEIRVGVOCIOKFLDIAROTCFPLQRLQGL
SVQFQVKEVSEJRNELQKQALVQKFLKSHWQCVLEIDINVQKRPADIPQGS
LAYEGASAKIPAPKPT"
BASE COUNT      410 a   24 c   309 g   399 t
ORIGIN
Query Match      88.2%   Score 19.4; DB 9; Length 1463;
Best Local Similarity 95.2%   Pred. No. 20;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

CY 2 ACCATGGCGGCTCCACTAGG 22
|||||
DB 10 AACATGGCGGCTCCACTAGG 30

RESULT 1:
AF121617 Homo sapiens (human)
LOCUS AF121617 Homo sapiens unknown mRNA 3109 bp mRNA linear PRI 07-DEC-2000
DEFINITION Homo sapiens unknown mRNA.
ACCESSION AF121617
VERSION AF121617.1 GI:11596417
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 3109)
AUTHOR Nguyen,M.H. and Liu,C.
TITLE Full-length cDNA for Unknown#3
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1409)
AUTHOR Nguyen,M.H. and Liu,C.
TITLE Direct Submission
JOURNAL Submitted (15-NOV-2000) Surgical Oncology, University of
California, Los Angeles, 10433 Le Centre Ave., Los Angeles, CA
90095, CA
FEATURES
Source Location/Qualifiers
1..3109
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1759..2295
/codon_start=1
/product="unknown"
/protein_id="AAK32724.1"
/db_xref="GI:13605460"
/translation="MAAPLGWVFSCQRPQPCAPPGLRQASHLQAAFGAPRPSSSTL
VDELSFACASVSCDYVNTQDEIRVGVOCIOKFLDIAROTCFPLQRLQGL
SVQFQVKEVSEJRNELQKQALVQKFLKSHWQCVLEIDINVQKRPADIPQGS
LAYEGASAKIPAPKPT"
BASE COUNT      980 a   503 c   621 g   1005 t
ORIGIN
Query Match      88.2%   Score 19.4; DB 9; Length 3109;
Best Local Similarity 95.2%   Pred. No. 35;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

CY 2 ACCATGGCGGCTCCACTAGG 22
|||||
DB 1756 AACATGGCGGCTCCACTAGG 1776

RESULT 12
AC006160 Homo sapiens chromosome 4 clone C0481F14 map 4p16, complete
LOCUS AC006160
DEFINITION Homo sapiens chromosome 4 clone C0481F14 map 4p16, complete
sequence.

```

```

ACCESSION AC006160 GI:5701616
VERSION AC006160.9
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 179310)
AUTHORS Stone,N.E., Schmutz,J.J., Shang,J., Pennacchio,L.A., Cox,D.R. and
Myers,R.M.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 179310)
AUTHORS Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
JOURNAL Direct Submission
COMMENT Submitted (28-DEC-1999) Department of Genetics, Stanford Human
Genome Center, 595 California Avenue, Palo Alto, CA 94304, USA
Quality: Phrap Quality >40 99.9% of Sequence.
Estimated Total Number of Errors is 0.1.
STS Content:
SHGC:36184 G30090
WT:13901 G21191
SHGC:50339 G33480
SHGC:50545 G33619
WT:15951 G24340
SHGC:79088 G48789
Location/Qualifiers
1..179310
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="4"
/map="4p16"
/clone="C0481F14"
/clone_lib="ROSWELL PARK CANCER RPCI - 11 Human Male BAC
Library"
BASE COUNT 50020 a 40057 c 39508 g 49725 t
ORIGIN
Query Match      88.2%   Score 19.4; DB 9; Length 179310;
Best Local Similarity 95.2%   Pred. No. 20;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

CY 2 AACATGGCGGCTCCACTAGG 20
|||||
DB 13622 AACATGGCGGCTCCACTAGG 93242

RESULT 13
AC11376/c Rattus norvegicus clone CH210-12204, *** SEQUENCING IN PROGRESS
LOCUS AC11376
DEFINITION Rattus norvegicus clone CH210-12204, *** SEQUENCING IN PROGRESS
*** 21 unordered pieces.
ACCESSION AC11376
VERSION AC11376.4 GI:30578929
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 278093)
AUTHORS Muzny,D., Metzker,M., Abamzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Arguiano,D.,
Ayalebechi,V., Ayagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,

```



```

misc_feature    complement(4435..5911)
                /note="Clone boundary
clone_end=77
size=EcoRI
end_sequence=58204243"
misc_feature    complement(234799..1235
                /note="Clone boundary
clone_end=506

```

	BASE COUNT	COUNT PERCENTAGE
ORIGIN		
79095	a 53766 c 54182 g	19149 : 11962 others

Quarry Match	20,000	Store 19,500	Length 2,000,000
Best local similarity	90,000	Prod. No. 139	
Matches	20	Conservative	20
		Mis-matches	20
		Indels	20
		Gars	0

29 : CACCAAGGGGGCTTCACCTAGGG 22

19163 CCCATGGCGTCTTCACTAGGG 19142

RESULT 14	
AC12545549	LOCUS
AC125455	DEFINITION
MUS MUSCULUS CHROMOSOME 10R CLON 4 RP1-7.7-12, WORKING CRAFT	SEQUENCE
AC125455	LOCUS
107893 bp	DNA
10648	MAP
FEB 28 JUN 2002	

ACCESSION AC125455  
VERSION AC125455.2 G1162870  
KEYWORDS HTS; HGS-PHASE1; HGS-DEPT.; HGS-FULLTOP  
SOURCE MUS MUSCULUS (HOUSE MOUSE)

**REFERENCES**

BUS MASCARDUS  
CROWTHER  
Eukaryota; Metazoa; Cnidaria; Vertebata; Euteleostomi;  
Mammalia; Euthera; Rodentia; Sturnonathii; Yuridae; Meridae; Mus.  
Bases 170-17293.

JOURNAL	unpublished
REFERENCE	2 (bases 1 to 17293)
AUTHORS	McPetersen, J.D. and Waterston, E.H.
TITLE	Direct Submission
JOURNAL	Submitted (26-CHN-2202) Genome Sequencing Center, 444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE	3 (bases 1 to 17293)
AUTHORS	McPetersen, J.D. and Waterston, E.H.
TITLE	Direct Submission
JOURNAL	Submitted (28-CHN-2000) Genome Sequencing Center, 444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT	On Jun 24, 2004 this sequence was added to the database.

Center: Washington University George Sengstack Center  
 Center code: WUGSC  
 Web site: <http://george.wustl.edu/georg/index.shtml>  
 Contact: [submissions@wustl.edu](mailto:submissions@wustl.edu)  
 Project Information  
 Center project name: M. Grottel

..... Summary Statistics

Sequencing vector: M16, 01

Sequencing primer: 5'-TTCCTG-3'

Chemistry: Dye primer, 0.6 reads

Assembly: Dye terminator, 318, 0.0% of reads

Assembly quality: 177.3, 0.0% of reads

Consensus quality: 177.3, 0.0% of reads

Consensus: 177.3, 0.0% of reads

- NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. The true order of the pieces

• is not known and their order in this sequence record is  
• arbitrary. Gaps between the contigs are represented as  
• runs of N, but the exact sizes of the gaps are unknown.  
• This record will be updated with the finished sequence  
• as soon as it is available and the accession number will  
• be preserved.

• 1 177291: contig of 177291 bp in length.

• 1 177293: contig of 177293 bp in length.

```

FEATURES             SOURCE
locationQualifiers  1..27723
                    /organism="Mus musculus"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:10090"
                    /chromosome="XNX"
                    /clone="EP24.17612"

```

```
misc feature
    17:293
    /note="assembly name:Contigs"
    BASE COUNT = 15:27 C 3380 G 5346 T
    ATGCN
```

```
Query Match          91.6%; Score 18; Length 17729;
Best Loca: Similarity 100.0%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0
```

QY 1 CACCATGGGCGCTCCACT 18

1:5:32 CACCATGGCGCTCCACT 1:5:15

RESULT : 5  
AC: 08948

LOCUS	AC108948	156921 bp	DNA	linear	RCO 2: JUN-2002
DEFINITION	Mus musculus clone PP23-183f12, complete sequence.				

NO: 55604 AC: 08948

VEES:CN	AC106948.1	GI:21536184
KEYWORDS	HTG.	
SOURCE	Mus musculus	house mouse
ORGANISM	Mus musculus	

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;  
Mus (Phase) (Phase)

REFERENCE  
McPherson J. D. and Waterston, R. H.  
1992. The sequence of *Mus musculus* clone  
2 (bases 1 to 194923)  
Unpublished

TITLE	Direct Submission
JOURNAL	Submitted (31-FEB-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE	3 (bases 1 to 196923)
AUTHORS	McPherson, J.D. and Waterston, R.H.
TIME	Direct Submission
JOURNAL	Submitted (21-FEB-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE	4 (bases 1 to 126923)
AUTHORS	McPherson, J.D. and Waterston, R.H.
TITLE	Direct Submission
JOURNAL	Submitted (21-JUN-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT	On Jun 21, 2002 this genome version replaced 019874255

Center: Washington University Genome Sequencing Center  
Center code: WJGSC  
Web site: <http://genome.wustl.edu/gsc/index.shtml>  
Contact: [submissions@wustl.wustl.edu](mailto:submissions@wustl.wustl.edu)  
Project information:  
Center project name: MGA0181312

```
FEATURES
Source
Location/Qualifiers
i: 196923
/organism="Mus musculus"
/mo_type="genomic DNA"
/acc_xref="taxon:10090"
```



```
BASE COUNT      59929 a 37584 c 38911 g 60499 t
ORIGIN
Query Match      81.8% Score 187 DB 102 Length 196923
Best Local Similarity 100.0% Pred.No. 1:1e+22
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY      1 CACCATGGCGGCTCCACT 18
DB      117538 CACCATGGCGGCTCCACT 117535
```

```
Search completed: October 20, 2003, 23:50:08
Job time : 124.403 secs
```

GenCore version: 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

ON nucleic nucleic search, using sw model  
Run on: October 20, 2003, 10:12:14 : Search time 10.3994 Seconds  
(without alignment)  
4672.369 Million cell updates/sec

Title: US 10 029-137 9

Perfect score: 18

Sequence: : rccaggttgct-cagagg 16

Scoring table: IDENTITY\_NUC

Gapop 10.0 : Gapext 1.0

Searched: 2552/56 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 2

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_190.mdb  
1: /SIDSI/gcgdata/geneseq/geneseq-emb./NA1989.DAT  
2: /SIDSI/gcgdata/geneseq/geneseq-emb./NA1981.DAT  
3: /SIDSI/gcgdata/geneseq/geneseq-emb./NA1982.DAT  
4: /SIDSI/gcgdata/geneseq/geneseq-emb./NA1983.DAT  
5: /SIDSI/gcgdata/geneseq/geneseq-emb./NA1984.DAT  
6: /SIDSI/gcgdata/geneseq/geneseq-emb./NA1985.DAT  
7: /SIDSI/gcgdata/geneseq/geneseq-emb./NA1986.DAT  
8: /SIDSI/gcgdata/geneseq/geneseq-emb./NA1987.DAT  
9: /SIDSI/gcgdata/geneseq/geneseq-emb./NA1988.DAT  
10: /SIDSI/gcgdata/geneseq/geneseq-emb./NA1989.DAT  
11: /SIDSI/gcgdata/geneseq/geneseq-emb./NA1990.DAT  
12: /SIDSI/gcgdata/geneseq/geneseq-emb./NA1991.DAT  
13: /SIDSI/gcgdata/geneseq/geneseq-emb./NA1992.DAT  
14: /SIDSI/gcgdata/geneseq/geneseq-emb./NA1993.DAT  
15: /SIDSI/gcgdata/geneseq/geneseq-emb./NA1994.DAT  
16: /SIDSI/gcgdata/geneseq/geneseq-emb./NA1995.DAT  
17: /SIDSI/gcgdata/geneseq/geneseq-emb./NA1996.DAT  
18: /SIDSI/gcgdata/geneseq/geneseq-emb./NA1997.DAT  
19: /SIDSI/gcgdata/geneseq/geneseq-emb./NA1998.DAT  
20: /SIDSI/gcgdata/geneseq/geneseq-emb./NA1999.DAT  
21: /SIDSI/gcgdata/geneseq/geneseq-emb./NA2000.DAT  
22: /SIDSI/gcgdata/geneseq/geneseq-emb./NA2001.DAT  
23: /SIDSI/gcgdata/geneseq/geneseq-emb./NA2002.DAT  
24: /SIDSI/gcgdata/geneseq/geneseq-emb./NA2003.DAT  
25: /SIDSI/gcgdata/geneseq/geneseq-emb./NA2004.DAT

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY

Result No.	Score	Query Match	Length DB	ID	Description
1	18	100.0	216	22	ABA46557 Human breast cell
2	18	100.0	216	22	ABA46420 Human foetal liver
3	18	100.0	216	22	AAK12876 Human brain expro
4	18	100.0	216	22	AAK18607 Human bone marrow
5	18	100.0	216	22	AA119407 Probe #3340 for ge
6	18	100.0	216	22	AA144555 Probe #3281 used t
7	18	100.0	216	22	AA105132 Probe #5223 used t
8	18	100.0	216	23	AB538171 Human liver single

9	18	100.0	216	24	AB512673 Human genome-deriv
10	18	100.0	478	22	AB51741 Human foetal liver
11	18	100.0	478	22	AAK05050 Human brain expro
12	18	100.0	478	22	AAK25487 Human bone marrow
13	18	100.0	478	22	AA110111 Probe #44 for gene
14	18	100.0	478	22	AA131357 Probe #41 used to
15	18	100.0	478	22	AA100552 Probe #43 used to
16	18	100.0	478	23	AB525055 Human liver single
17	18	100.0	478	24	AB525053 Human genome-deriv
18	18	100.0	952	20	AAZ77452 Human ovarian tumo
19	18	100.0	1071	20	AAZ33600 Human breast tumo
20	18	100.0	1941	23	AA576383 DNA encoding novel
21	18	100.0	10395	25	AA374584 Secreted protein g
22	17	94.4	540	22	AA583986 Mouse Dpi coding s
23	17	94.4	1002	22	AA583986 Mouse Dpi coding s
24	17	94.4	1425	22	AA583986 Mouse Dpi coding s
25	16.4	91.1	1544	23	AA578292 DNA encoding novel
26	16.4	91.1	1544	23	AA578292 DNA encoding novel
27	16.4	91.1	1544	23	AA578292 DNA encoding novel
28	15.4	85.6	508	22	AAK65665 Human immune/haema
29	15.4	85.6	1080	20	AAZ00410 Human secreted pro
30	15.4	85.6	1752	24	AB583812 Human prostate exp
31	15.4	85.6	5147	23	ABV25206 DNA encoding novel
32	15	83.3	1126	23	AA567517 DNA encoding novel
33	15	83.3	1173	23	AA567517 DNA encoding novel
34	15	83.3	1173	23	AA567517 DNA encoding novel
35	14.8	82.2	204	18	AA188802 Nuclear steroid ho
36	14.8	82.2	307	21	AA411345 Human secreted exp
37	14.8	82.2	323	22	AAK69720 Human immune/haema
38	14.8	82.2	325	22	AAK69720 Human immune/haema
39	14.8	82.2	396	24	AB59609 Novel murine polyn
40	14.8	82.2	484	22	AAK89392 Barley clone F. H
41	14.8	82.2	576	22	AAK03225 Human cDNA clone i
42	14.8	82.2	578	24	AB578472 Human ovarian cano
43	14.8	82.2	909	22	AA527437 Human apoptosis-as
44	14.8	82.2	915	23	AA571708 DNA encoding novel
45	14.8	82.2	1042	22	AA160747 Human polynucleoti

ALIGNMENT'S

RESULT 1  
ABA46557  
ID ABA46557 standard; DNA; 216 BP.  
XX ABA46557  
AC ABA46557  
DT 01-FEP-2002 first entry;  
XX  
XX Human breast cell single exon nucleic acid probe #5242.  
DE  
XX Human: Microarray; single exon probe; gene expression; breast;  
XX disease; cancer; ss.  
XX  
XX Homo sapiens.  
XX  
XX NC200157271-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30 JAN-2001; 2001WO-USC0662.  
XX  
XX 04-FEB-2000; 2000US-0180112.  
XX 26-MAY-2000; 2000US-0207456.  
XX 30-JUN-2000; 2000US-0608408.  
XX 03-AUG-2000; 2000US-0632366.  
XX 21-SEP-2000; 2000US-0234687.  
XX 27-SEP-2000; 2000US-0236359.  
XX 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MODE-1) MOLECULAR DYNAMICS INC.  
XX

```

PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-496933/54.
XX
XX New spatially-addressable set of single exon nucleic acid probes,
XX useful for measuring gene expression in sample derived from human
XX breast, comprises number of single exon nucleic acid probes -
XX
XX Claim 4: SEQ ID NO 5252; 327pp + sequence listing; English.
XX
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human breast and B1 474 cells. The method involves contacting
XX the probes with a collection of detectably labeled nucleic acids
XX derived from RNA of human breast, and then measuring the label
XX bound to each probe of the microarray. The probes are useful for:
XX verifying the expression of regions of genomic DNA predicted to
XX encode proteins. They are useful for gene discovery, and for
XX determining predisposition and/or prognosing breast disease. Gene
XX expression analysis is useful for assessing the toxicity of chemical
XX agents on cells. The microarray of this invention presents a far greater
XX diversity of probes for measuring gene expression, with far less bias
XX than expressed sequence tag microarrays. The method is suitable for:
XX rapid production of functional information from genomic sequence. The
XX present sequence is a single exon nucleic acid probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WFO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 216 BP; 39 A; 55 C; 60 G; 62 T; 0 other;
SQ

```

Query Match: 100.0%; Score 18; DB 22; Length 216;  
Best Local Similarity: 100.0%; Pred. No. 8.7;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 TCACGTTGGCTTCAGAG 18
    |||||
DB 1 TCACGTTGGCTTCAGAG 18

RESULT 2
ABA64420
ID ABA64420 standard; DNA; 216 BP.
AC ABA64420;
XX
XX 01-FEB-2002 (first entry)
XX
XX Human foetal liver single exon nucleic acid probe #12725.
XX
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
XX Homo sapiens.
XX
XX WO200157277 A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-050669.
XX
XX 04-FEB-2002; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0628408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-024263.
XX
XX (MOLE-1) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX
XX Example 4: SEQ ID NO: 12569; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,

```

Human genome derived single exon nucleic acid probes useful for  
analyzing gene expression in human fetal liver -  
Claim 4: SEQ ID NO 12725; 639pp + sequence listing; English.  
The invention relates to a single exon nucleic acid probe for  
measuring human gene expression in a sample derived from human foetal  
liver. The single exon nucleic acid probes may be used for predicting,  
measuring and displaying gene expression in samples derived from human  
fetal liver. The present sequence is a single exon nucleic acid  
probe of the invention.  
Note: The sequence data for this patent did not form part of the  
printed specification, but was obtained in electronic format directly  
from WFO at ftp.wipo.int/pub/published\_pct\_sequences.  
Sequence 216 BP; 39 A; 55 C; 60 G; 62 T; 0 other;

Query Match: 100.0%; Score 18; DB 22; Length 216;  
Best Local Similarity: 100.0%; Pred. No. 8.7;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACGTTGGCTTCAGAG 18  
|||||  
DB 1 TCACGTTGGCTTCAGAG 18

RESULT 3  
AAK12878  
ID AAK12878 standard; DNA; 216 BP.  
XX  
XX AAK12878;  
XX  
XX 05-NOV-2001 (first entry)  
XX  
XX Human brain expressed single exon probe SEQ ID NO: 12569.  
XX  
XX Human; brain expressed exon; gene expression analysis; probe;  
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
XX epilepsy; cancer; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200157275 A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-050667.  
XX  
XX 04-FEB-2002; 2000US-0180312.  
XX  
XX 26-MAY-2000; 2000US-0207456.  
XX  
XX 30-JUN-2000; 2000US-0628408.  
XX  
XX 03-AUG-2000; 2000US-0632366.  
XX  
XX 21-SEP-2000; 2000US-0234687.  
XX  
XX 27-SEP-2000; 2000US-0236359.  
XX  
XX 04-OCT-2000; 2000GB-024263.  
XX  
XX (MOLE-1) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483447/52  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human  
XX brains -  
XX  
XX Example 4: SEQ ID NO: 12569; 650pp + Sequence Listing; English.  
XX  
XX The present invention provides a number of single exon nucleic acid  
XX probes which are derived from genomic sequences expressed in the human  
XX brain. They can be used to measure gene expression in brain cell samples,  
XX which may enable the diagnosis and improved treatment of nervous system  
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,

```
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
SQ Sequence 216 BP; 39 A; 55 C; 60 G; 62 T; 0 other;

Query Match      100.0%; Score 18; DB 22; Length 216;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCACGTTGGCTTCAGAGG 18
    |||||
Db 1 TCACGTTGGCTTCAGAGG 18

RESULT 4
AAK19407
ID AAK19407 standard; DNA; 216 BP.
XX
AC AAK19407;
XX
CT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 13164.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; yeloma; ss.
XX
OS Homo sapiens.
XX
PN WC200157276-A2.
XX
PC 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00659.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0074263.
XX
PA (MOLR) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WP1; 2001-488901/53.
XX
DR Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow.
XX
PS Claim 25; SEQ ID NO 9340; 487bp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SNPs). The present sequence is one such probe. The SNPs are derived
CC from human HeLa cells. The SNPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: the sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPC
CC at http://wipo.int/pub/published/pat\_sequences.
XX
SQ Sequence 216 BP; 39 A; 55 C; 60 G; 62 T; 0 other;

Query Match      100.0%; Score 18; DB 22; Length 216;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCACGTTGGCTTCAGAGG 18
    |||||
Db 1 TCACGTTGGCTTCAGAGG 18

RESULT 6
AAI44595
ID AAI44595 standard; DNA; 216 BP.
XX
AC AAI44595;
XX
CT 17-OCT-2001 (first entry)
XX
DE Probe #1328 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
```

```
AAI19407
ID AAI19407 standard; DNA; 216 BP.
XX
AC AAI19407;
XX
CT 12-OCT-2001 (first entry)
XX
DE Probe #9340 for gene expression analysis in human cervical cell sample.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WC200157276-A2.
XX
PC 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00659.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0074263.
XX
PA (MOLR) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WP1; 2001-488901/53.
XX
DR Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells
XX
PS Claim 25; SEQ ID NO 9340; 487bp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SNPs). The present sequence is one such probe. The SNPs are derived
CC from human HeLa cells. The SNPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: the sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPC
CC at http://wipo.int/pub/published/pat\_sequences.
XX
SQ Sequence 216 BP; 39 A; 55 C; 60 G; 62 T; 0 other;

Query Match      100.0%; Score 18; DB 22; Length 216;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCACGTTGGCTTCAGAGG 18
    |||||
Db 1 TCACGTTGGCTTCAGAGG 18

RESULT 6
AAI44595
ID AAI44595 standard; DNA; 216 BP.
XX
AC AAI44595;
XX
CT 17-OCT-2001 (first entry)
XX
DE Probe #1328 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
```

```

XX OS Homo sapiens.
XX PN WO200157273 A2.
XX PC 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00661.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608458.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE): MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR.
XX DR WPI; 2001-455937/53.
XX CC Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human placenta
XX PS Claim 25; SEQ ID No 13281; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SNPs).
XX CC The present sequence is one such probe. The probes are useful for
XX CC producing a microarray for predicting, measuring and displaying gene
XX CC expression in samples derived from human placenta. The probes are useful
XX CC for antenatal diagnosis of human genetic disorders.
XX SQ Sequence 216 BP; 39 A; 55 C; 60 G; 62 T; 0 other.

Query Match 100.0%; Score 18; DB 22; Length 216;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1 TCACCTTGCTTCAGAGG 18
DB : TCACCTTGCTTCAGAGG 18

RESULT 2
AA105112 standard; CNA; 216 BP.
XX AC AA105112;
XX DT 09-OCT-2001 (first entry)
XX DE Probe #5123 used to measure gene expression in human breast sample.
XX KW Probe; human; breast disease; breast cancer; development disorder; ser;
XX KW inflammatory disease; proliferative breast disease; non-melanoma tumor.
XX OS Homo sapiens
XX PN WO200157273-A2.
XX PC 09-AUG-2001.
XX PF 29-JAN-2001; 2001WO-US00661.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.

```

```

XX PA (MOLE): MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR.
XX DR WPI; 2001-476286/41.
XX CC Novel single exon nucleic acid probe used to measuring gene expression
XX PT in a human breast.
XX PS Claim 25; SEQ ID No 6123; 172pp; English.
XX CC The present invention relates to novel single exon nucleic acid probes.
XX CC The present sequence is one such probe. The probes are useful for
XX CC measuring human gene expression in a human breast sample, where the probe
XX CC hybridizes at high efficiency to a nucleic acid expressed in the human
XX CC breast. The probes are useful for predicting, measuring, displaying,
XX CC storing, analyzing and producing gene expression data from tissues.
XX CC Particularly, the probes are useful for predicting, measuring, displaying
XX CC of the breast, with gene expression, at relative breast disease and
XX CC non-carcinoma models.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic form directly from WPI
XX CC at 15P-WPI-001(pub)/published_seq_sequences.
XX SQ Sequence 216 BP; 39 A; 60 C; 60 G; 62 T; 0 other.

Query Match 100.0%; Score 18; DB 22; Length 216;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1 TCACCTTGCTTCAGAGG 18
DB : TCACCTTGCTTCAGAGG 18

RESULT 8
ARS38171
XX CD ARS38171 standard; CNA; 216 BP.
XX AC ARS38171;
XX DT 25-FEB-2003 (first entry)
XX DE Human liver; single exon probe; SEQ ID No 1161.
XX KW Human; single exon nucleic acid probe; liver; cirrhosis;
XX KW hyperlipoproteinemia; hyperlipidaemia; hypercholesterolemia;
XX KW coronary heart disease; is.
XX OS Homo sapiens
XX PN WO200157273-A2.
XX PC 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US-00661.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE): MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR.
XX DR WPI; 2001-489898/53.

```

PT Human genome derived single exon nucleic acid probes useful for  
 PT analysing gene expression in human adult liver;  
 XX  
 XX Claim 4: SEQ ID NO 13161; 698pp; English.  
 XX  
 CC The invention relates to a single exon nucleic acid probe (SENPA) for  
 CC measuring human gene expression in a sample derived from human adult  
 CC liver, comprising one of 13109 defined nucleotide sequences given in the  
 CC specification for components/ fragments). The probe hybridises at high  
 CC stringency to a nucleic acid molecule expressed in the human adult  
 CC liver. (i) may be used for predicting, measuring and displaying gene  
 CC expression in samples derived from human adult liver. The genes  
 CC identified may be involved in genetic liver diseases such as cirrhosis,  
 CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which  
 CC is associated with coronary heart disease. ABS5111, ABS5112 represent  
 CC human liver single exon nucleic acid probes of the invention.  
 CC Note: The sequence information for this patent does not appear in the  
 CC printed specification but was obtained in electronic format directly  
 CC from WIPO at [http://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).  
 XX  
 SQ Sequence 216 BP; 19 A; 55 C; 60 G; 62 T; 0 other;  
 Query Match 100.0%; Score 18; PB 24; Length 216;  
 Best Local Similarity 100.0%; Pred. No. 8.7;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TCACGTTGGCTTCAGAG 18  
 DE 1 TCACGTTGGCTTCAGAG 18  
 RESULT 9  
 ABS12673  
 ID ABS12673 standard; DNA: 216 BP.  
 XX  
 AC ABS12673;  
 XX  
 CT 19-ABS-2002 (first entry)  
 XX  
 DE Human genome-derived single exon probe GAP from lung SEQ ID NO 12664.  
 XX  
 KW Human ds; single exon probe; asthma; lung cancer; COPD; emphysema;  
 KW chronic obstructive pulmonary disease; interstitial lung disease;  
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KW tuberculosis; Gaucher's disease; Niemann-Pick disease;  
 KW Hermannsky-Pudlak syndrome; sarcoidosis; pulmonary haemorrhoidosis;  
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagazer syndrome;  
 KW primary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KW primary ciliary dyskinesia; pulmonary hypertension;  
 KW hyaline membrane disease; open breathing frame; 44  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200106003 A2.  
 XX  
 PD 15-NOV-2001.  
 XX  
 PF 10-JAN-2001; 2001WO-0060455.  
 XX  
 FR 04-FEB-2003; 2000US-180312P.  
 PR 26-MAY-2003; 2000US-207456P.  
 PR 30-JUN-2003; 2000US-0608428.  
 PR 03-AUG-2003; 2000US-0612166.  
 PR 21-SEP-2003; 2000US-234587P.  
 PR 27-SEP-2003; 2000US-236359P.  
 PR 04-OCT-2003; 2000GB-0024263.  
 XX  
 (MOLE : MOLECULAR DYNAMICS INT.  
 PA Penn SG, Hanzel DK, Chen W, Rank DR;  
 PI WFT; 2002-114183/15.  
 XX  
 XX

PT Spatially-addressable set of single exon nucleic acid probes, used to  
 PT measure gene expression in human lung samples;  
 XX  
 XX Claim 4: SEQ ID NO 12664; 634pp; English.  
 XX  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 1387 open reading frames derived from the 12614  
 CC probes. Also included are a microarray comprising the novel set of  
 CC probes. The novel set of probes which hybridise at high stringency to a  
 CC nucleic acid expressed in the human lung, comprising (a) contacting the array with  
 CC a sample derived from human lung, comprising (a) contacting the array with  
 CC a collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA, and (b) measuring the label detectably bound to each probe of  
 CC the array; identifying exons in a eukaryotic genome, comprising  
 CC (a) algorithmically predicting at least one exon from genomic sequences  
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably  
 CC labeled nucleic acids from eukaryotic lung mRNA to a single exon probe,  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above-mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridisation to a single exon  
 CC microarrays having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types indicates that  
 CC the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene  
 CC expression analysis, and for identifying exons in a gene, particularly  
 CC using human lung derived mRNA and for the study of lung diseases  
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease  
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,  
 CC Niemann-Pick disease, Hermannsky-Pudlak syndrome, lymphangioleiomyomatosis,  
 CC haemorrhoidosis, pulmonary histiocytosis, Karagazer syndrome, fibrocystic  
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
 CC and hyaline membrane disease. The present sequence is a single exon  
 CC probe open reading frame of the invention.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC [http://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).  
 XX  
 SQ Sequence 216 BP; 19 A; 55 C; 60 G; 62 T; 0 other;  
 Query Match 100.0%; Score 18; PB 24; Length 216;  
 Best Local Similarity 100.0%; Pred. No. 8.7;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TCACGTTGGCTTCAGAG 18  
 DE 1 TCACGTTGGCTTCAGAG 18  
 RESULT 10  
 ABS1741  
 ID ABS1741 standard; DNA: 478 BP.  
 XX  
 AC ABS1741;  
 XX  
 CT 01-FEB-2002 (first entry)  
 XX  
 DE Human foetal liver single exon nucleic acid probe #46.  
 XX  
 KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.  
 OS Homo sapiens.  
 XX  
 PN W0200107277 A2.  
 XX

```

PD 09-AUG-2001.
XX (MOLE-) MOLECULAR DYNAMICS INC.
PF Penn SG, Hanzel PK, Chen W, Rank DR.
XX WPI: 2001-481447/52.
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel PK, Chen W, Rank DR.
XX WPI: 2001-481447/52.
XX Human genome derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver.
XX Claim 1: SEQ ID NO 46; 639pp + sequence listing; English.
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human fetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pat_sequences.
XX Sequence 475 BP; 136 A; 111 C; 100 G; 211 T; 0 other;
SQ
Query Match 100.0%; Score 18; PB 22; Length 475;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TCAAGTTGGCTTCAGAGG 18
DB 270 TCAAGTTGGCTTCAGAGG 287
RESULT 1:
AAK00050
ID AAK00050 standard; DNA; 475 BP.
XX AAK00050;
XX 05-NOV-2001 (first entry);
XX Human brain expressed single exon probe SEQ ID NO: 41.
DE Human brain expressed single exon probe SEQ ID NO: 41.
XX Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX Homo sapiens.
XX WO200157275-A2.
PN 09-AUG-2001.
XX 09-AUG-2001.
XX 10-JAN-2001; 2001WO-US05667.
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.

```

```

XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel PK, Chen W, Rank DR.
XX WPI: 2001-481447/52.
XX Single exon nucleic acid probes for analyzing gene expression in human
PT brains.
XX Example 4: SEQ ID NO: 41; 639pp + sequence listing; English.
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and autism. The present sequence is one of the probes of the
CC invention.
XX Sequence 475 BP; 136 A; 111 C; 100 G; 211 T; 0 other;
SQ
Query Match 100.0%; Score 18; PB 22; Length 475;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TCAAGTTGGCTTCAGAGG 18
DB 270 TCAAGTTGGCTTCAGAGG 287
RESULT 12
AAK25467
ID AAK25467 standard; DNA; 475 BP.
XX AAK25467;
XX 06-NOV-2001 (first entry);
XX Human bone marrow expressed single exon probe SEQ ID NO: 44.
XX Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukemia; lymphoma; myeloma; ss.
XX Homo sapiens.
XX WO200157276-A2.
PN 09-AUG-2001.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US05668.
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel PK, Chen W, Rank DR.
XX WPI: 2001-481447/53.
XX Human genome derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow.
XX Example 4: SEQ ID NO: 44; 658pp + sequence listing; English.
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human

```

CC bone marrow. They can be used to measure gene expression in bone marrow  
 CC samples, which may enable the improved diagnosis and treatment of cancers  
 CC such as lymphoma, leukemia and myeloma. The present sequence is one of  
 CC the probes of the invention.

XX  
 SQ Sequence 478 BP; 136 A; 111 C; 100 G; 131 T; 0 Other;

Query Match 100.0%; Score 18; DP 22; Length 478;  
 Best Local Similarity 100.0%; Pred. No. 9.3;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACGTGGCTTCACAGG 18  
 |||||  
 Db 270 TCACGTGGCTTCACAGG 287

RESULT 14  
 AA131357  
 ID AA131357 standard; DNA; 478 BP.

XX  
 AC AA131357;

XX  
 DT 12-OCT-2001 (first entry)

XX  
 DE Probe #44 for gene expression analysis in human cervical cell sample.

XX  
 KW Probe; human; microarray; gene expression; cervical epithelial cells;  
 KW cervical cancer; ss.

XX  
 CS Homo sapiens.

XX  
 PN W0200157272-A2.

XX  
 PD 09-AUG-2001.

XX  
 PF 30-JAN-2001; 2001WO-US006670.

XX  
 PR 04-FEB-2002; 2000US-0180312.

XX  
 PR 26-MAY-2002; 2000US-0207456.

XX  
 PR 30-JUN-2002; 2000US-0603409.

XX  
 PR 03-AUG-2002; 2000US-0612366.

XX  
 PR 21-SEP-2002; 2000US-0234687.

XX  
 PR 27-SEP-2000; 2000US-0236359.

XX  
 PR 04-OCT-2000; 2000GB-0024263.

XX  
 PA (MOLFE) MOLECULAR DYNAMICS INC.

XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX  
 WP1; 2001-488907/53.

XX  
 PT Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human cervical epithelial cells

XX  
 PS Claim 25; SEQ ID No 44; 487bp; English.

XX  
 CC The present invention relates to human single exon nucleic acid probes  
 CC (SENPs). The present sequence is one such probe. The SENPs are derived  
 CC from human Hela cells. The SENPs can be used to produce a single exon  
 CC microarray, which can be used for measuring human gene expression in a  
 CC sample derived from human cervical epithelial cells. By measuring gene  
 CC expression, the probes are therefore useful in grading and/or staging  
 CC of diseases of the cervix, notably cervical cancer.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic form: directly from WIPO  
 CC at [http://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

XX  
 SQ Sequence 478 BP; 136 A; 111 C; 100 G; 131 T; 0 Other;

Query Match 100.0%; Score 18; DP 22; Length 478;  
 Best Local Similarity 100.0%; Pred. No. 9.3;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACGTGGCTTCACAGG 18  
 |||||  
 Db 270 TCACGTGGCTTCACAGG 287

RESULT 14  
 AA131357

ID AA131357 standard; DNA; 478 BP.

XX  
 AC AA131357;

XX  
 DT 12-OCT-2001 (first entry)

XX  
 DE Probe #44 used to measure gene expression in human placenta sample.

XX  
 KW Probe; microarray; human; placenta; antenatal diagnosis;  
 KW genetic disorder; ss.

XX  
 CS Homo sapiens.

XX  
 PN W0200157272-A2.

XX  
 PD 09-AUG-2001.

XX  
 PF 30-JAN-2001; 2001WO-US006670.

XX  
 PR 04-FEB-2002; 2000US-0180312.

XX  
 PR 26-MAY-2002; 2000US-0207456.

XX  
 PR 30-JUN-2002; 2000US-0603409.

XX  
 PR 03-AUG-2002; 2000US-0612366.

XX  
 PR 21-SEP-2002; 2000US-0234687.

XX  
 PR 27-SEP-2000; 2000US-0236359.

XX  
 PR 04-OCT-2000; 2000GB-0024263.

XX  
 PA (MOLFE) MOLECULAR DYNAMICS INC.

XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX  
 WP1; 2001-488907/53.

XX  
 PT Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human placenta -

XX  
 PS Claim 25; SEQ ID No 43; 654bp; English.

XX  
 CC The present invention relates to single exon nucleic acid probes (SENPs).  
 CC The present sequence is one such probe. The probes are useful for  
 CC producing a microarray for predicting, measuring and displaying gene  
 CC expression in samples derived from human placenta. The probes are useful  
 CC for antenatal diagnosis of human genetic disorders.

XX  
 SQ Sequence 478 BP; 136 A; 111 C; 100 G; 131 T; 0 Other;

Query Match 100.0%; Score 18; DP 22; Length 478;

Best Local Similarity 100.0%; Pred. No. 9.3;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACGTGGCTTCACAGG 18

|||||

Db 270 TCACGTGGCTTCACAGG 287

RESULT 15

AA100052

ID AA100052 standard; DNA; 478 BP.

XX  
 AC AA100052;

XX  
 DT 09-OCT-2001 (first entry)

XX  
 DE Probe #43 used to measure gene expression in human breast sample.

XX  
 KW Probe; human; breast disease; breast cancer; development disorder; ss;



KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.

XX Homo sapiens.

XX WO200157270-A2.

XX PD 09-AUG-2001.

XX PR 29-JAN-2001; 25C:WO-US00661.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000SB-0024263.

XX PA (MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanze DK, Chen W, Rank DR;

XX DR WPI; 2001-476286/51.

XX PT Novel single exon nucleic acid probe used to measuring gene expression

XX PT in a human breast -

XX PS Claim 25; SEQ ID No 43; 122pp; English.

XX CC The present invention relates to novel single exon nucleic acid probes.

XX CC The present sequence is one such probe. The probes are useful for

XX CC measuring human gene expression in a human breast sample, where the probe

XX CC hybridises at high stringency to a nucleic acid expressed in the human

XX CC breast. The probes are useful for predicting, diagnosing, grading,

XX CC staging, monitoring and prognosing diseases of the human breast,

XX CC particularly those diseases with polygenic aetiology. The diseases

XX CC include: breast cancer, disorders of development, inflammatory diseases

XX CC of the breast, fibrocystic changes, proliferative breast disease and

XX CC non-carcinoma tumours.

XX CC Note: The sequence data for this patent did not form part of the printed

XX CC specification, but was obtained in electronic format directly from Wipo

XX CC at ftp.wipo.int/pub/published\_pat\_sequences.

XX SQ Sequence 478 BP; 136 A; 111 C; 100 G; 131 T; 0 other;

Query Match 100.0%; Score 18; DB 22; Length 478;

Best Local Similarity 100.0%; Pred. No 9.3;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACGTTGGCTTCAGAGG 18

Db 270 TCACGTTGGCTTCAGAGG 287

Search completed: October 21, 2003, 00:03:56

Job time : 11.3934 secs

GenCore version 5.1.6  
Copyright (c) 1995 - 2003 CompuGen Ltd.

OV nucleic nucleic search, using sw model

Run on: October 20, 2003, 18:30:49 Search time 100.460 seconds

(without alignments)  
4354.325 Million cell updates/sec

Filter: US-10-029-137-9

Perfect score: 18

Sequence: 1:caagtgagatcaagag 18

Scoring table: IDENTITY NC

Gapop 10.0, lapext 1.0

Searched: 22761392 seqs, 1215223856 residues

Total number of hits satisfying chosen parameters: 45562734

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: EST

1: e-estba\*

2: e-esthu\*

3: e-estma\*

4: e-estma\*

5: e-estor\*

6: e-estor\*

7: e-estor\*

8: e-estor\*

9: e-estor\*

10: g-est2\*

11: g-est2\*

12: g-est2\*

13: g-est2\*

14: g-est2\*

15: e-estma\*

16: e-estma\*

17: e-estma\*

18: e-estma\*

19: e-estma\*

20: e-estma\*

21: e-estma\*

22: e-estma\*

23: e-estma\*

24: e-estma\*

25: e-estma\*

26: e-estma\*

27: e-estma\*

28: g-est2\*

29: g-est2\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	100.0	251	9 A1539573	A1539573 (p15510.x
2	18	100.0	276	12 BQ211961	BQ211961 U1-1-BC1p
3	18	100.0	279	12 BQ211918	BQ211918 U1-1-BC1p
4	18	100.0	365	9 A1539577	A1539577 (p15510.x

C	5	18	100.0	382	14 CA338172
C	6	18	100.0	423	14 W88616
C	7	18	100.0	420	14 W99361
C	8	18	100.0	417	9 AA339643
C	9	18	100.0	425	9 AA339643
C	10	18	100.0	437	9 AA339643
C	11	18	100.0	437	9 AA339643
C	12	18	100.0	444	9 AA339643
C	13	18	100.0	456	9 AA339643
C	14	18	100.0	461	9 AA339643
C	15	18	100.0	465	9 AA339643
C	16	18	100.0	465	9 AA339643
C	17	18	100.0	465	9 AA339643
C	18	18	100.0	463	9 AA339643
C	19	18	100.0	463	9 AA339643
C	20	18	100.0	477	9 AA339643
C	21	18	100.0	482	14 W99361
C	22	18	100.0	487	13 BX282583
C	23	18	100.0	489	14 W67773
C	24	18	100.0	500	9 AA331022
C	25	18	100.0	526	13 BU071817
C	26	18	100.0	532	9 AW503116
C	27	18	100.0	536	9 A1932315
C	28	18	100.0	537	9 A1128801
C	29	18	100.0	542	9 AW351929
C	30	18	100.0	549	9 BF439680
C	31	18	100.0	557	12 BQ211920
C	32	18	100.0	559	9 A1929343
C	33	18	100.0	559	9 AW734731
C	34	18	100.0	562	10 BF439687
C	35	18	100.0	563	10 BE937115
C	36	18	100.0	567	14 CB165119
C	37	18	100.0	569	12 BQ211920
C	38	18	100.0	570	12 BQ211920
C	39	18	100.0	573	14 CA952729
C	40	18	100.0	573	14 CA150523
C	41	18	100.0	581	14 CA337193
C	42	18	100.0	591	12 BQ211920
C	43	18	100.0	593	12 BQ211920
C	44	18	100.0	595	12 BQ211920
C	45	18	100.0	617	12 BM272234
C	46	18	100.0	621	14 CA952484

ALIGNMENTS

A1539573 251 bp tRNA tRNA  
T1539573: NCI-CGAP 104 Homo sapiens cDNA clone IMAGE1539573,  
similar to tRNA and repetitive element; contains TARI TARI  
TARI repetitive element 1, tRNA sequence.

A1539573 GI:4453708  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-ram@nci.nih.gov](mailto:cgaps-ram@nci.nih.gov)

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CGA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Gene distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:

www-bio.lnln.gov/bbrp/image/image.html  
 Insert Length: 683 Std Error: 0.00  
 Seq Primer: 40bp from Gibco  
 High quality sequence stop: 203  
 POLYA=NC.

## FEATURES

source

## Location/Qualifiers

1..251  
 /organism="Homo sapiens"  
 /mol\_type="rRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2189955"  
 /tissue\_type="serous papillary carcinoma, high grade, 2 pooled tumors"  
 /lab\_hosts="CHICB"  
 /clone\_lib="NCI\_CGAP\_Ut4"  
 /note="Organ: Ovary; Vector: pCMV-SPOPT6; Site: 1; Salt: Site 2; Not 1; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.48 kb. Life Technologies catalog #: 11342-316"  
 46 3 91 C 70 g 44 t

## BASE COUNT

ORIGIN

Query Match 100.0% Score 18; DB 9; Length 251;  
 Best Local Similarity 100.0%; Pred. No. 9;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QV 1 TCACGTTGGCTTCAGAG 14

||||| ||||| |||||

74 TCACGTTGGCTTCAGAG 91

## RESULT 2

B0011981

LOCUS

DEFINITION UI-1-BC1p-atg-a-06-C U1-1 NCI CGAP P13 Homo sapiens cDNA clone  
 UI-1-BC1p-atg-a-06-C-U1-1, mRNA sequence.

ACCESSION

B0011981

VERSION

B0011981.1

KEYWORDS

EST

SOURCE

Homo sapiens

ORGANISM

Homo sapiens (human);

REFERENCE

1..276

AUTHORS

NCI CGAP

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

JOURNAL

Unpublished

COMMENT

Contact: Robert Strausberg, Ph.D.  
 Email: rgs@bbs.fda.gov  
 Tissue Procurement: Dr. Steven Brown  
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 cDNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, bento-soares@uiowa.edu  
 The following repetitive elements were found in this cDNA  
 sequence: 1-25, >AT richlow\_complexity  
 Seq primer: M13 FORWARD  
 POLYA=Yes

## FEATURES

source

## Location/Qualifiers

1..276  
 /organism="Homo sapiens"  
 /mol\_type="rRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-1-BC1p-atg-a-06-C-U1-1"  
 /tissue\_type="placenta"  
 /dev\_stages="9-9 weeks"  
 /lab\_hosts="DHICB (Life Technologies)"  
 /clone\_lib="NCI\_CGAP\_P.3"  
 /note="Organ: Placenta; Vector: pT7T3-Pac (Pharmacia); with a modified polylinker; Site 1: EcoR I; Site 2: Not I;  
 NCI\_CGAP\_P13 is a subtracted cDNA library constructed

according to Bethesda, Lennan and Soares, Genome Research, 4:461-476, 1996. First strand cDNA synthesis was primed with a 30 bp primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The original recipe used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are GA, AAGAA, For additional information, contact: Bento Soares, bento-soares@uiowa.edu

TAG: 1-25, >AT richlow\_complexity  
 TAG: 1-25, >AT richlow\_complexity  
 TAG: 1-25, >AT richlow\_complexity

BASE COUNT 67 3 41 1 1 Others  
 ORIGIN

Query Match 100.0% Score 18; DB 9; Length 251;  
 Best Local Similarity 100.0%; Pred. No. 9;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QV 1 TCACGTTGGCTTCAGAG 14

||||| ||||| |||||

74 TCACGTTGGCTTCAGAG 91

## RESULT 3

B0011981

LOCUS

DEFINITION UI-1-BC1p-atg-a-06-C U1-1 NCI CGAP P13 Homo sapiens cDNA clone  
 UI-1-BC1p-atg-a-06-C-U1-1, mRNA sequence.

ACCESSION

B0011981

VERSION

B0011981.1

KEYWORDS

EST

SOURCE

Homo sapiens

ORGANISM

Homo sapiens (human);

REFERENCE

1..276

AUTHORS

NCI CGAP

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

JOURNAL

Unpublished

COMMENT

Contact: Robert Strausberg, Ph.D.  
 Email: rgs@bbs.fda.gov  
 Tissue Procurement: Dr. Steven Brown  
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 cDNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, bento-soares@uiowa.edu  
 The following repetitive elements were found in this cDNA  
 sequence: 1-25, >AT richlow\_complexity  
 Seq primer: M13 FORWARD  
 POLYA=Yes

## FEATURES

source

## Location/Qualifiers

1..279  
 /organism="Homo sapiens"  
 /mol\_type="rRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-1-BC1p-atg-a-06-C-U1-1"  
 /tissue\_type="Placenta"  
 /dev\_stages="9-9 weeks"  
 /lab\_hosts="DHICB (Life Technologies)"  
 /clone\_lib="NCI\_CGAP\_P13"  
 /note="Organ: Placenta; Vector: pT7T3-Pac (Pharmacia); with a modified polylinker; Site 1: EcoR I; Site 2: Not I;  
 NCI\_CGAP\_P13 is a subtracted cDNA library constructed according to Bethesda, Lennan and Soares, Genome Research, 4:461-476, 1996. First strand cDNA synthesis was primed with a 30 bp primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac

vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the 3' tail. The sequence tags for this library are GA, AGGAA. For additional information, contact: Bento Soares.

Bento-Soares@uoi.br  
TAG\_Lib=U1-1-RC1p  
TAG\_Tissue=placenta human 8 week  
TAG\_SEQ=GA

BASE COUNT  
ORIGIN

Query Match 100.0%; Score 18; DB 12; Length 279;  
Best Local Similarity 100.0%; Pred. No. 1.1e-02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACGTTGGCTTCAGAG 18  
|||||

DB 172 TCACGTTGGCTTCAGAG 189

RESULT 4

LOCUS

DEFINITION A1358977 365 bp mRNA linear EST 15 FEB 1999  
Q19901.X1 NCI-CCAP\_Bin23 Homo sapiens cDNA clone IMAGE:2012496 3';  
mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human);

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo;

1 (bases 1 to 365)

NCI/NINDS-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute / National Institute of Neurological

Disorders and Stroke, Brain Tumor Genome Anatomy Project

CCGAP/BRCAP; Tumor Gene Index

Unpublished

Contact: Robert Strausberg, Ph.D.

E-mail: cgapbs@mail.nih.gov

Tissue Procurement: David N. Louis, M.D., Nynke R. Rosenfeld M.D.,

Ph.D.

CNA Library Preparation: M. Bento Soares, Ph.D., M. Marina

Bonafido, Ph.D.

cDNA Sequencing by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CCAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www.biology.washington.edu/llnl

Insert length: 365 Std Error: 0.0

Seq primer: 40bp from 5' end

Location/Qualifiers

1..365

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/cclone="IMAGE:2012496"

/tissue\_type="glioblastoma (poorly"

/lab\_host="DH10B"

/clone\_lib="NCI-CCAP Bin23"

/note="Organ: Brain; Vector: pTZ19 Lac (Pharmacia) with a

modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer (5'

GTGTACATCTGAGTGGAGCGCGGATCTTTTCTTTTCTTTTCTTTT

3'); double-stranded cDNA was ligated to Eco RI

adaptors (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of the modified pTZ19 vector.

Library is normalized, and was constructed by Bento

Soares and Mafatima Bonafido."

118 a 75 c 78 g 94 t

BASE COUNT  
ORIGIN

Query Match 100.0%; Score 18; DB 9; Length 365;  
Best Local Similarity 100.0%; Pred. No. 1.1e-02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACGTTGGCTTCAGAG 18  
|||||

DB 276 TCACGTTGGCTTCAGAG 293

RESULT 5

LOCUS

DEFINITION CA338172 362 bp mRNA linear EST 04 NOV 2002  
NISC\_L1W12a06.Y: COGENE 4PA; Homo sapiens cDNA clone IMAGE:560971  
5'; mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo;

1 (bases 1 to 362)

NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP);

Tumor Gene Index

Unpublished

Contact: Robert Strausberg, Ph.D.

E-mail: cgapbs@mail.nih.gov

CNA Library Preparation:

DNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL

Sequencing Center (NISC)

Clone distribution: NCI-CCAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

info@image.llnl.gov

Plate: J2AM12415 row: B column: 12

Seq primer: M13RP1 reverse primer (AR1).

Location/Qualifiers

1..362

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/cclone="IMAGE:560971"

/tissue\_type="1st pharyngeal arch (poorly"

/dev\_stage="embryo, 4 weeks postconception"

/lab\_host="DH10B"

/clone\_lib="COGENE 4PA"

/note="Vector: pAMP1; cDNA primed using oligo-dT primer,

directionally cloned into UD3 sites of pAMP1. Size

selected for insert sizes ranging from 0.2-1.2 kb.

Normalized to 600.0. Primary library, non-amplified.

Library constructed by M. Lovett. For more information on

this library, please contact R. Tidwell (Washington

University) or visit the COGENE website at

http://hg.wustl.edu/COGENE/."

113 a 72 c 75 g 122 t

BASE COUNT  
ORIGIN

Query Match 100.0%; Score 18; DB 14; Length 382;  
Best Local Similarity 100.0%; Pred. No. 1.1e-02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACGTTGGCTTCAGAG 18  
|||||

DB 61 TCACGTTGGCTTCAGAG 44

RESULT 6

LOCUS

DEFINITION W88616 403 bp mRNA linear EST 07-MAY-1997  
zh71d04.s1 Soares fetal liver spleen; NFILS\_1; Homo sapiens cDNA  
clone IMAGE:417511 3'; mRNA sequence.

W88616

Accession

```

VERSION W88616.1 GI:1404136
SOURCE EST.
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 403)
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, R.,
Chissee, S., Dietrich, N., Dubuque, T., Faveillo, A., Gish, W., Hawkins
M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore
B., Morris, M., Parsons, C., Prange, C., Rifkin, L., Rohlfing, T.,
Schellenberg, K., Soares, M.B., Tan, F., Thierky-Meg, J., Trevas, E.,
Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Maria, M.
The WashU-Merck EST Project
TITLE Generation and analysis of 283,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996);
MEDLINE 97244478
PUBMED 8889549
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8001, St. Louis, MO 63128
Tel: 314 286 1828
Fax: 314 286 1818
Email: est@watson.wustl.edu
This clone is available royalty free through LNC. ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 649 Std Error: 0.00
Seq Primer: mob R5GA-RT
High quality sequence stop: 317.
Location/Qualifiers
source
1..403
/organism="Homo sapiens"
/molecule="mRNA"
/db_xref="Gene:1325981"
/db_xref="taxon:9606"
/clone="IMAGE:417511"
/sex="male"
/dev_stage="20 week-post conception fetus"
/ab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares fetal heart NHEH-9W Homo sapiens cDNA clone
with a modified polylinker; Site 1: Pac 1; Site 2: Eco RI;
This is a subtracted version of the original Soares fetal
liver spleen INFES library. 1st strand cDNA was primed
with a Pac 1 oligo(dT) primer [5].
AATGGGAAGATTAAATTAAGATCTTTTCTTTTCTTTT 3'.
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac 1 and cloned into the Pac 1
and Eco RI sites of the modified p773 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 93 a 99 c 92 g 118 t
ORIGIN
Query Match: 100.0%; Score 18; DB 14; Length 403;
Best Loc: Similarity 100.0%; Pred. No. 1.1e-02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Cy 1 TCACGTTGGGCTTCAGAGG 18
Db 150 TCACGTTGGGCTTCAGAGG 167

RESULT 7
W99361
LOCUS z001f08.s1 Soares fetal heart NHEH-9W Homo sapiens cDNA clone
DEFINITION IMAGE:357927 3', mRNA sequence.
ACCESSION W99361 GI:1435245
VERSION W99361.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 417)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Maria, M., Parsons, C.,
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevas, E., Waterston
R., Williamson, A., Wohlmann, P. and Wilson, R.
The WashU-Merck EST Project
TITLE Unpublished
JOURNAL Contact: Wilson RK
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 417)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Maria, M., Parsons, C.,
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevas, E., Waterston
R., Williamson, A., Wohlmann, P. and Wilson, R.
The WashU-Merck EST Project
TITLE Unpublished
JOURNAL Contact: Wilson RK

```

Washington University School of Medicine  
 4444 Forest Park Parkway, Box 501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: estewatson.wustl.edu  
 This clone is available royalty free through LNCX; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -29M13 rev2 from Amersham  
 High quality sequence stop: 315.  
 Location/Qualifiers  
 .. 427

#### FEATURES

Source  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="G38:1284571"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:376315"  
 /sex="unknown"  
 /dev\_stage="19 weeks"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /clone\_lib="Soares fetal heart NBH19W"  
 /notes="Organ: heart; Vector: pTV3D (Pharmacia) with a  
 modified polylinker; Site 1: Not 1; Site 2: Eco RI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5'  
 TGTACACATCGAGTGGACGCGCCGACCTTTTCTTTTCTT 3',  
 double-stranded cDNA was size selected, ligated to Eco RI  
 adapters (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of a modified pTV3 vector  
 (Pharmacia). Library went through one round of  
 normalization to a Cot = 5. Library constructed by  
 M. Fatima Bonaldo. This library was constructed from the  
 same fetus as the fetal lung library, Soares fetal lung  
 NBH19W."  
 102 a 55 c 88 g 141 t 1 others

#### BASE COUNT

ORIGIN

Query Match 100.0%; Score 18; DB 9; Length 417;  
 Best Local Similarity 100.0%; Pred. No. 1:2e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 TCACGTGGCTTCAGAG 18

Db 102 TCACGTGGCTTCAGAG 95

#### RESULT 9

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

#### TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

Initial assessment of human Gene diversity and expression patterns  
 based upon 83 million nucleotides of cDNA sequence  
 Nature 377 (6547 Suppl), 3-174 (1995)  
 96026280  
 7566098  
 Other ESTs: EST:2622 THCI74495  
 Contact: Kerlavage, AR  
 Bioinformatics  
 The Institute for Genomic Research  
 9712 Medical Center Drive, Rockville, MD 20850 USA  
 Tel: 3018699056  
 Fax: 3018699423  
 Email: arkerlav@igtr.org

For clone availability, additional sequence and expression  
 information related to this EST, please check the TIGR Human Gene  
 Index (<http://www.tigr.org/db/hgi/hgi.htm>)  
 Seq primer: M13 21

#### FEATURES

Source

Location/Qualifiers

.. 425

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="ATCC:192954"

/db\_xref="taxon:9606"

/dev\_stage="adult"

/clone\_lib="uterus tumor 1"

/notes="Organ: uterus; Vector: pBluescript SK+ Site 1:  
 EcoRI; Site 2: XhoI"

BASE COUNT 126 a 95 c 90 g 112 t 2 others

ORIGIN

Query Match 100.0%; Score 18; DB 9; Length 425;  
 Best Local Similarity 100.0%; Pred. No. 1:2e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 TCACGTGGCTTCAGAG 19

Db 275 TCACGTGGCTTCAGAG 292

#### RESULT 10

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AA812894 437 bp mRNA linear EST 31-DEC-1998  
 ai81f03.s: Soares, testis, NHT: Homo sapiens cDNA clone 317245 3,  
 mRNA sequence.

AA812894 1 GI:2882958  
 EST.

Homo sapiens (human)  
 Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 437)  
 NCICGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index

Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgabs@mail.nih.gov](mailto:cgabs@mail.nih.gov)

Ph.D.  
 DNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo

cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center

Clone distributor: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:

[www.bic.llnl.gov/bbrp/image/image.html](http://www.bic.llnl.gov/bbrp/image/image.html)  
 Insert length: 1527 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 427.

Location/Qualifiers  
 .. 437

/organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"



```

DE      336 TCACGTTGGCTTCAGAGG 343
      ||| ||||| ||||| |||||
RESULT 13
LOCUS   A1818636          456 bp     mRNA     linear     EST 07-MAR-2000
DEFINITION WK89a06.X1 NCI_CGAP_Kids Homo sapiens cDNA clone IMAGE:2422546 3',
            mRNA sequence.
ACCESSION A1818636
VERSION   A1818636.1 G1:1437215
KEYWORDS EST, Homo sapiens (human)
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 456)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap/
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE    Tumor Gene Index
JOURNAL  Unpublished
COMMENT  Contact: Robert Strausberg, Ph.D.
        Email: cgapbs@mail.nih.gov
        Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
        Emmert-Buck, M.D., Ph.D.
        cDNA Library Preparation: M. Bento Soares, Ph.D.
        cDNA Library Arrayed by: Greg Lennon, Ph.D.
        DNA Sequencing by: Washington University Genome Sequencing Center
        Clone distribution: NCI-CGAP clone distribution information can be
        found through the I.M.A.G.E. Consortium/LNL at:
        www.bio.llnl.gov/bbrip/image/image.html
        Insert length: 499 Std Error: 0.00
        Seq primer: -40BP from 5' end
        High quality sequence stop: 444.
        Location/Qualifiers
            1..456
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:2422546"
            /tissue_type="squamous cell carcinoma, poorly
            differentiated (4 pooled tumors, including primary and
            metastatic)"
            /dev_stage="adult"
            /lab_host="NCI CGAP Kids"
            /note="Organ: Lung; Vector: pMT23-Pac (Pharmacia) with a
            modified polylinker; 1st strand cDNA was prepared from
            pooled lung tumor tissue, and was then primed with a Not I
            and R1 adaptors (Pharmacia), digested with Not I and
            Eco RI sites of the modified pMT23 vector. Library was
            constructed by Bento Soares and M. Fatima Bonaldo (Soares4
            1)."
            117 a 107 c 98 g 134 t
BASE COUNT 117 a 107 c 98 g 134 t
ORIGIN
Query Match 100.0%; Score 18; DB 9; Length 456;
Best local Similarity 100.0%; Pred No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1 TCACGTTGGCTTCAGAGG 18
    ||| ||||| |||||
DB 157 TCACGTTGGCTTCAGAGG 174
    ||| ||||| |||||

RESULT 15
LOCUS   A1344731          461 bp     mRNA     linear     EST 02-FEB-1999
DEFINITION qp05a06.X1 NCI_CGAP_Kids Homo sapiens cDNA clone IMAGE:1917106 3',
            mRNA sequence.
ACCESSION A1344731
VERSION   A1344731.1 G1:4081937
KEYWORDS EST, Homo sapiens (human)
SOURCE   Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 461)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap/
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE    Tumor Gene Index
JOURNAL  Unpublished
COMMENT  Contact: Robert Strausberg, Ph.D.
        Email: cgapbs@mail.nih.gov

```



Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 cDNA Library Arrayed by: Greg Lennan, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the L.M.A.G.E. Consortium/HLN at: [www.biclini.gov/bicp/image/image.html](http://www.biclini.gov/bicp/image/image.html)  
 Insert length: 784 Std Error: 0.00  
 Seq primer: 4TUP from Gibco  
 High quality sequence stop: 413.

FEATURES

SOURCE

```

1..465
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1917106"
/tissue_type="2 pooled tumors (clear cell type)"
/lab_host="DH10B"
/clone_idb="NCI-CGAP_Kids"
/notes="Organ: Kidney; Vector: pMT0 Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(18) primer (5' AACTGGAAGAAATTCGGCGCCGCAATCTTTTCTTTTCTTTT 3'), double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pMT0 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."
129 a 128 c 96 g 112 t

```

BASE COUNT  
 ORIGIN

Query Match: 100.0%; Score 18; DH 9; Length 465;  
 Best Local Similarity 100.0%; Pred. NC: 1.2e12;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TCACGTTGGTTCAGAGG 18  
 |||||  
 Db 257 TCACGTTGGTTCAGAGG 274

Search completed: October 21, 2003, 02:09:46  
 Job time : 101.456 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

XX nucleic nucleic search, using sw model  
Run on: October 20, 2003, 17:12:14 ; Search time 97.6934 seconds  
(without alignments)  
7537.592 Million cell updates/sec  
Title: US-10-029-137-9  
Perfect score: 18  
Sequence: 1 tgaacgttggttcagagatg  
Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0  
Searched: 2888711 seqs, 2345481386 residues  
Total number of hits satisfying chosen parameters: 5777422  
Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenBank

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fur.\*
- 17: em.hu.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ph.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.bur.\*
- 31: em.htg.invi.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.roi.\*
- 36: em.htg.mar.\*
- 37: em.htg.vit.\*
- 38: em.sy.\*
- 39: em.htgo.fur.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by change to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	18	100.0	789	9	AF317680
C 2	18	100.0	822	6	BC015361
C 3	18	100.0	863	6	AX014817
C 4	18	100.0	1069	9	AF317679
C 5	18	100.0	1071	6	AX017517
C 6	18	100.0	1071	6	BC014527
C 7	18	100.0	1093	9	BC011936
C 8	18	100.0	1088	6	AX015050
C 9	18	100.0	1363	9	AF358829
C 10	18	100.0	3103	9	AF31627
C 11	18	100.0	179310	9	AC006160
C 12	17	94.4	1002	6	AX099133
C 13	17	94.4	1719	10	AF165166
C 14	17	94.4	1929	10	BC025140
C 15	17	94.4	1933	10	AF165165
C 16	17	94.4	1955	10	AF192385
C 17	17	94.4	2149	10	AF192383
C 18	17	94.4	3192	10	AF192384
C 19	17	94.4	3386	10	AF192382
C 20	17	94.4	35760	10	U29187
C 21	17	94.4	183681	2	AC137814
C 22	17	94.4	206549	10	AL833794
C 23	17	94.4	28359	2	AC139886
C 24	17	94.4	250290	2	AC121081
C 25	16.4	91.1	587	11	BV066209
C 26	16.4	91.1	1057	1	AB085490
C 27	16.4	91.1	1058	1	AB085431
C 28	16.4	91.1	43050	6	AX695476
C 29	16.4	91.1	55956	2	AC131559
C 30	16.4	91.1	85304	9	HS227L5
C 31	16.4	91.1	119561	9	AL138799
C 32	16.4	91.1	142542	2	AC134538
C 33	16.4	91.1	148851	9	HS155D22
C 34	16.4	91.1	163456	9	AC017071
C 35	16.4	91.1	179612	2	AC126274
C 36	16.4	91.1	179612	2	AC126274
C 37	16.4	91.1	231559	2	AC104882
C 38	16.4	91.1	234239	10	AC114819
C 39	16.4	91.1	235091	2	AC130530
C 40	16.4	91.1	231914	2	AC087566
C 41	16.4	91.1	271233	2	AC109029
C 42	16	88.9	10384	1	AF004205
C 43	16	88.9	156089	2	AC017350
C 44	15.4	85.6	254	11	BV005958
C 45	15.4	85.6	255	11	BV006424

ALIGNMENTS

RESULT 1  
AF317680/c  
LOCUS AF317680 Homo sapiens tumor-related protein mRNA, complete cds.  
DEFINITION AF317680  
ACCESSION AF317680  
VERSION AF317680.1 GI:13021849  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 789)  
AUTHORS  
TITLE  
A gene from tumor  
JOURNAL

AF317680 789 bp mRNA linear PRI 21-FEB-2001

```

REFERENCE 2 (bases 1 to 789)
AUTHORS Mai,N. and Lib,C.
TITLE Direct Submission
JOURNAL Submitted (01-NOV-2003) Surgical Oncology, University of
California, Los Angeles, 10833 Le Conte Ave., Los Angeles, CA
90095, USA

FEATURES
Source
1..789
/organism="Homo sapiens"
/mol_type="cDNA"
/db_xref="taxon:9606"
21..593
/codon_start=1
/product="tumor-related protein"
/protein_id="AAK1563.1"
/db_xref="GI:13021850"
/translation="MAAPLGMPGSGGPGPPAPPLPGASLLQAAAGAPRPSSTL
VDESSPACAFSLVQDYNGTQEEIEITGQDQICAFLEIAROTGEFELCARLQ
VQESPEQV:KEVSELSNE:QRKDALVCNKL:KLAWQGVLEDTNVCIRKKA:IPQSS
LAV:GAGICGHPITSEANVSGRQ:AYEWALA"
216 a 179 c 194 g 198 t 2 others
BASE COUNT 216 a 179 c 194 g 198 t
ORIGIN
1 TCACGTTGGCTTCAGAG 18
|||||
549 TCACGTTGGCTTCAGAG 512

Query Match 100.0%; Score 18; DB 9; Length 789;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACGTTGGCTTCAGAG 18
|||||
549 TCACGTTGGCTTCAGAG 512
Db

RESULT 2
BD015361/c
LOCUS Human protein and cDNA[4] 822 bp DNA linear PAT 27-AUG-2002
DEFINITION BD015361
ACCESSION BD015361
VERSION BD015361.1 GI:22556499
KEYWORDS JP 2001218584-A/7
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 822)
Human protein and cDNA[4]
Patent: JP 2001218584-A 7 14-AUG-2002;
JAPAN SCIENCE AND TECHNOLOGY CORP
OS Homo sapiens (human)
PN JP 2001218584-A/7
PC 14-AUG-2001
PF 28-FEB-2000 JP 2000031062
PI SEIJI KATO,MIHORO SAKKI
PC C12N15/09,C07K14/435,C07K16/18,C12N1/15,C12N1/19,C12N1/21,PC
C12N5/107/
FC C12P21/38,C12M15/00,C12M5/80
CC
FH Key Location/Qualifiers
FT CDS location:Qualifiers
1..822
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
210 a 186 c 199 g 227 t
BASE COUNT 210 a 186 c 199 g 227 t
ORIGIN

Query Match 100.0%; Score 16; DB 6; Length 822;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACGTTGGCTTCAGAG 18
|||||
549 TCACGTTGGCTTCAGAG 512
Db

REFERENCE 2 (bases 1 to 789)
AUTHORS Mai,N. and Lib,C.
TITLE Direct Submission
JOURNAL Submitted (01-NOV-2003) Surgical Oncology, University of
California, Los Angeles, 10833 Le Conte Ave., Los Angeles, CA
90095, USA

FEATURES
Source
1..789
/organism="Homo sapiens"
/mol_type="cDNA"
/db_xref="taxon:9606"
21..593
/codon_start=1
/product="tumor-related protein"
/protein_id="AAK1563.1"
/db_xref="GI:13021850"
/translation="MAAPLGMPGSGGPGPPAPPLPGASLLQAAAGAPRPSSTL
VDESSPACAFSLVQDYNGTQEEIEITGQDQICAFLEIAROTGEFELCARLQ
VQESPEQV:KEVSELSNE:QRKDALVCNKL:KLAWQGVLEDTNVCIRKKA:IPQSS
LAV:GAGICGHPITSEANVSGRQ:AYEWALA"
216 a 179 c 194 g 198 t 2 others
BASE COUNT 216 a 179 c 194 g 198 t
ORIGIN
1 TCACGTTGGCTTCAGAG 18
|||||
549 TCACGTTGGCTTCAGAG 512

Query Match 100.0%; Score 18; DB 9; Length 789;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACGTTGGCTTCAGAG 18
|||||
549 TCACGTTGGCTTCAGAG 512
Db

RESULT 3
AX014817/c
LOCUS Sequence 3 from Patent WO990049 882 bp DNA linear PAT 07-SEP-2000
DEFINITION AX014817
ACCESSION AX014817
VERSION AX014817.1 GI:10541054
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 882)
Human protein and cDNA[4]
Patent: WO990049 A 23-SEP-2000;
SCHMITZ ARVID FOR NUTRITIONALS INC FOR THE FARM EDCAP GENE; HILTMANN
HEIDI GENE; KENTHAL AMERSON LTD MEDICAL GENE FUIP GENOMICS/HUN
DELL PHILIPPEY CHRISTIAN GENE
FEATURES
Source
1..882
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
218 a 180 c 206 g 248 t
BASE COUNT 218 a 180 c 206 g 248 t
ORIGIN

Query Match 100.0%; Score 18; DB 9; Length 882;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACGTTGGCTTCAGAG 18
|||||
549 TCACGTTGGCTTCAGAG 512
Db

RESULT 4
AF117679/c
LOCUS Homo sapiens tumor related cDNA sequence 1069 bp mRNA linear PAT 27-FEB-2001
DEFINITION AF117679
ACCESSION AF117679
VERSION AF117679.1 GI:13021847
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1069)
A gene from tumor
Published
2 (bases 1 to 1069)
Mat.N. and Lib.C.
Direct Submission
TITLE Direct Submission
JOURNAL Submitted (01-NOV-2000) Surgical Oncology, University of
California, Los Angeles, 10833 Le Conte Ave., Los Angeles, CA
90095, USA
FEATURES
Source
1..1069
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..1069
/gene="tumor-related"
BASE COUNT 213 a 216 c 256 g 105 t 19 others
ORIGIN

Query Match 100.0%; Score 18; DB 9; Length 1069;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACGTTGGCTTCAGAG 18
|||||
549 TCACGTTGGCTTCAGAG 512
Db

```

```

CY 1 TCACGTGGCTTCAGAG 18
   |||||
Db 541 TCACGTGGCTTCAGAG 524

RESULT 5
AX017517/C
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
AUTHORS
Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Pilarczyk,C.
TITLE
Human nucleic acid sequences from normal breast tissue
JOURNAL
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METZGER GERT FUER GENOMFORSCHEN
(DE); PILARSKY CHRISTIAN (DE)
FEATURES
Location/Qualifiers
1..1071
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT
281 a 218 c 255 g 317 t
ORIGIN
Query Match 100.0% Score 18; DB 6; Length 1071.
Best local Similarity 100.0%; Pred. No. 21;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1 TCACGTGGCTTCAGAG 18
   |||||
Db 546 TCACGTGGCTTCAGAG 524

RESULT 7
BC011936/C
LOCUS
DEFINITION
Homo sapiens, hypothetical protein DKFZp434K153, clone WGC:20562
IMAGE:1936713, mRNA, complete cds.
ACCESSION
BC011936
VERSION
BC011936.1 GI:115780366
KEYWORDS
MGC
SOURCE
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
AUTHORS
Strausberg R.
TITLE
Direct Submission.
JOURNAL
Submitted 130-JUL-2001; National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
COMMENT
NCH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: mgc-help@nci.nih.gov
Tissue Procurement: DCC/DTP
CMA Library Preparation: Rubin Laboratory
CMA Library Arrayed by: The I.M.A.G.E. Consortium (ILND)
CMA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC)
Gaithersburg, Maryland
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgc@nih.nih.gov
Shevchenko,Y., Wetherby,K.D., Beckstrom-Sternberg,S.M.,
Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brickley,C., Brooks,S.,
Dietrich,N.L., Guan,X., Gupta,J., Ho,S.-L., Karlins,E., Legaspi,R.,
Lim,M., Maduro,Q.L., Masliah,C., Mastrian,S.O., McCloskey,J.C.,
McDowell,J., Pearson,R., Snyder,B., Stantrispop,S., Thomas,P.J.,
Thompson,B.L., Touchman,J.W., Tsurganov,C., Vogt,J.L., Walker,M.A.,
Zhang,L.H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILND at: http://image.llnl.gov
Series: IRAL Plate: 28 Row: f Column: 4
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis:
Location/Qualifiers
1..1023
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="LOCUS:80336"
/db_xref="taxon:9606"
/clone="WGC:20562 IMAGE:3936713"
/tissue_type="lung, small cell carcinoma"

```



SVQKPEGVKHFVSELRNELPKKDALVOKHITKLRLHWQVLELNVQHKRPAD:PCQS  
 CAYEGASANIPIALKE"

BASE COUNT 980 a 503 c 621 g 1605 t

ORIGIN

Query Match 100.0%; Score 18; DB 9; Length 1799;

Best Local Similarity 100.0%; Pred. No. 23;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACGTGGCTTCAGAG 14

DB 2295 TCACGTGGCTTCAGAG 2478

RESULT 11

AC006160

LOCUS

DEFINITION

AC006160

AC006160.9

HTG

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

```

CDS
46..585
/codon_start=1
/product="pion protein-like protein bpl"
/protein_id="AA02545.1"
/db_xref="GI:1926351"
/translation="MKNRIGTMMVAILOMLASHLSTVKARGIKRFRKWRKRVLPSSG
GQTEARVAENRPGAFIKQGRKLDIDFGAENRYVAANYWQPDGTYEGGSEANVTK
EMLVTSVNAIQANCAEFSEKQSKLHQRVLRWLKELCSAKHCDPFLERGAALV
AVQPAWCVLLSEVWIVK"
polyA_site
1214 438 c 434 g 415 t
BASE COUNT
412 a 438 c 434 g 415 t
ORIGIN
Query Match 94.4% Score 17; DB 10; Length 1759;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACGTTGGCTTACAG 17
|||||
Db 346 TCACGTTGGCTTACAG 143

RESULT 14
BC025140/c
LOCUS
DEFINITION Mus musculus, pion protein dublet, clone MGC:35610 IMAGE:360731,
mRNA, complete cds.
ACCESSION BC025140
VERSION BC025140.1 GI:1926351.5
KEYWORDS MGC
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1929)
AUTHORS Strausberg, R.
DIRECT SUBMISSION
Submitted (05-MAR-2003): National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://rgc.nci.nih.gov
Contact: MGC help desk
Email: cgapso@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amgacm.tmc.edu
Gunnar, P.H., Garcia, A.M., Lu, X., Huby, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: MAK Plate: 6; Row: d Column: 19
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 19265923.
FEATURES
Location/Qualifiers
1..1929
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:35610 IMAGE:360731"
/tissue_type="Mammary tumor, Meta:lochioren-TGF alpha
model: 10 month old virgin mouse. Taken by biopsy."
/clone_lib="NCI CGAP Xaml"
/lab_host="DH1CB"
/note="vector: pcMV-SPORT6"

CDS
1..1933
/organism="Mus musculus"
/mol_type="mRNA"
/strain="BALB/c"
/db_xref="taxon:10090"
/clone="19"
/tissue_type="tumor"
/db_xref="adult"
/contig="1933 downstream of pion protein locus;
alternatively spliced; contains exons 1b, 2 and 3"
42..193
/codon_start=1
/product="pion protein-like protein bpl"
/protein_id="AA02544.1"
/db_xref="GI:1604929"
/translation="MKNRIGTMMVAILOMLASHLSTVKARGIKRFRKWRKRVLPSSG
GQTEARVAENRPGAFIKQGRKLDIDFGAENRYVAANYWQPDGTYEGGSEANVTK
EMLVTSVNAIQANCAEFSEKQSKLHQRVLRWLKELCSAKHCDPFLERGAALV
AVQPAWCVLLSEVWIVK"
1919
polyA_site

```

BASE COUNT 491 a 478 c 465 g 433 t  
ORIGIN  
Query Match 94.4% Score 17 DB 10 Length 1932  
Best Local Similarity 100.0% Pred. No. 88  
Matches 17 Conservative 0 Mismatches 0 Indels 0 Gaps 0  
CY 1 TCACGTTGCCTTCAG 17  
DB 342 TCACGTTGCCTTCAG 346  
Search completed: October 21, 2003, 21:50:03  
Job time : 100.693 secs



Genore version 5.1.6  
Copyright (c) 1994-2003 CompuGen Ltd.

GM nucleic acid - nucleic search. us10-029-137-8

Run on: October 20, 2003, 21:40:34, Search time 2.414s, Seconds  
without alignment:  
2239.602 Million cell. of bases/sec

Title: US-10-029-137-8

Perfect score: 18

Sequence: 1: ATGGGCGGTCCTCACTAGGG 18

Scoring table: IDENTITY MB

Gapop 10.0, Gapext 1.0

Searches: 56978 seqs, 2269166 residues

Total number of hits satisfying chosen parameters: 1139356

Minimum 25 seq length: 6

Maximum 25 seq length: 200000000

Post processing: Minimum Match 31

Maximum Match 1000

Clustering first 45 summaries

Database: 1: /cgn2/6/prodata/2/na/na\_CGM seqs

2: /cgn2/6/prodata/2/na/na\_CGM seqs

3: /cgn2/6/prodata/2/na/na\_CGM seqs

4: /cgn2/6/prodata/2/na/na\_CGM seqs

5: /cgn2/6/prodata/2/na/na\_CGM seqs

6: /cgn2/6/prodata/2/na/na\_CGM seqs

Note: Rni is the number of results predicted by Rni to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY

Result No.	Score	Query	Length	DB	Description
1	14.8	80.0	48	US-09-300-958A-25	Sequence 25, Appl
2	14.8	80.0	48	US-09-300-958A-25	Sequence 25, Appl
3	14.8	80.0	48	US-09-300-958A-25	Sequence 25, Appl
4	14.8	80.0	48	US-09-300-958A-25	Sequence 25, Appl
5	14.8	80.0	48	US-09-300-958A-25	Sequence 25, Appl
6	14.8	80.0	48	US-09-300-958A-25	Sequence 25, Appl
7	14.8	80.0	48	US-09-300-958A-25	Sequence 25, Appl
8	14.8	80.0	48	US-09-300-958A-25	Sequence 25, Appl
9	13.8	76.7	48	US-09-300-958A-25	Sequence 25, Appl
10	13.8	76.7	48	US-09-300-958A-25	Sequence 25, Appl
11	13.8	76.7	48	US-09-300-958A-25	Sequence 25, Appl
12	13.8	76.7	48	US-09-300-958A-25	Sequence 25, Appl
13	13.8	76.7	48	US-09-300-958A-25	Sequence 25, Appl
14	13.8	76.7	48	US-09-300-958A-25	Sequence 25, Appl
15	13.8	76.7	48	US-09-300-958A-25	Sequence 25, Appl
16	13.8	76.7	48	US-09-300-958A-25	Sequence 25, Appl
17	13.8	76.7	48	US-09-300-958A-25	Sequence 25, Appl
18	13.8	76.7	48	US-09-300-958A-25	Sequence 25, Appl
19	13.8	76.7	48	US-09-300-958A-25	Sequence 25, Appl
20	13.8	76.7	48	US-09-300-958A-25	Sequence 25, Appl
21	13.8	76.7	48	US-09-300-958A-25	Sequence 25, Appl
22	13.8	76.7	48	US-09-300-958A-25	Sequence 25, Appl
23	13.8	76.7	48	US-09-300-958A-25	Sequence 25, Appl
24	13.8	76.7	48	US-09-300-958A-25	Sequence 25, Appl
25	13.8	76.7	48	US-09-300-958A-25	Sequence 25, Appl
26	13.8	76.7	48	US-09-300-958A-25	Sequence 25, Appl
27	13.8	76.7	48	US-09-300-958A-25	Sequence 25, Appl

28	13.4	74.4	393	4	US-09-252-991A-6435	Sequence 6435, Ap
29	13.4	74.4	414	4	US-09-252-991A-6356	Sequence 6356, Ap
30	13.4	74.4	597	4	US-09-252-991A-6120	Sequence 6120, Ap
31	13.4	74.4	693	4	US-09-252-991A-6194	Sequence 6194, Ap
32	13.4	74.4	845	2	US-08-773-910-2	Sequence 2, Appl
33	13.4	74.4	943	4	US-09-199-892-2	Sequence 2, Appl
34	13.4	74.4	1183	4	US-09-252-991A-5757	Sequence 5757, Ap
35	13.4	74.4	1616	3	US-08-818-112-112	Sequence 112, App
36	13.4	74.4	1616	4	US-08-818-111-107	Sequence 107, App
37	13.4	74.4	1616	4	US-09-056-556-112	Sequence 112, App
38	13.4	74.4	1616	4	US-09-272-586-107	Sequence 107, App
39	13.4	74.4	1616	4	US-09-996-243-263	Sequence 263, App
40	13.4	74.4	1779	4	US-09-252-991A-6267	Sequence 6267, App
41	13.4	74.4	1616	4	US-09-620-312D-439	Sequence 439, Appl
42	13.4	74.4	2361	3	US-08-705-771-7	Sequence 7, Appl
43	13.4	74.4	2382	4	US-09-167-109-1	Sequence 1, Appl
44	13.4	74.4	2382	5	ECT-US95-16980-2	Sequence 2, Appl
45	13.4	74.4	2640	4	US-09-252-991A-6033	Sequence 6033, Ap

ALIGNMENTS

RESULT 1  
US-09-300-958A-25  
Sequence 25, Application US/09100959A  
Patent No. 6495319  
GENERAL INFORMATION:  
APPLICANT: Nucleland, Michael  
APPLICANT: Welsh, John  
APPLICANT: Trekkie, Thomas  
TITLE OF INVENTION: Reduced Complexity Nucleic Acid Targets and Methods of Use  
FILE REFERENCE: PPH 1457  
CURRENT APPLICATION NUMBER: US/09/300-958A  
CURRENT FILING DATE: 1999-04-27  
PRIOR APPLICATION NUMBERS: 60/4081,331  
PRIOR FILING DATE: 1998-04-27  
PRIOR APPLICATION NUMBER: 60/095,070  
PRIOR FILING DATE: 1998-08-27  
PRIOR APPLICATION NUMBER: 60/118,634  
PRIOR FILING DATE: 1999-02-04  
NUMBER OF SEQ ID NOS: 85  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 25  
LENGTH: 48  
TYPE: DNA  
SEQUENCE: HMM Seq 100s  
FEATURES:  
NAME KEY: unsure  
LOCATION: 1000  
NAME KEY: unsure  
LOCATION: 1402  
NAME KEY: unsure  
LOCATION: 1459  
US-09-300-958A-25  
Query Match: 82.2%; Score 14.8; DB 4; length 463;  
Best Local Similarity: 88.9%;  
Matches: 16; Conservative: 0; Mismatches: 2; Indels: 6; Gaps: 6;  
CY 1 ATGGCGGTCCTCACTAGGG 18  
||| ||| ||| |||  
DB 137 ATGGCGGTCCTCACTAGGG 154  
RESULT 2  
US-08-462-1695-5/C  
Sequence 5, Application US/08462169B  
Patent No. 5792562  
GENERAL INFORMATION:  
APPLICANT: John Gruene and Craig A. Rosen  
TITLE OF INVENTION: Fictitious Growth Factor-15

```

? NUMBER OF SEQUENCES: 32
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
? ADDRESSEE: CECCHI, STEWART & OLSTEIN
? STREET: 6 BECKER FARM ROAD
? CITY: ROSELAND
? STATE: NEW JERSEY
? COUNTRY: USA
? ZIP: 07068
? COMPUTER READABLE FORM:
? MEDIUM TYPE: 3.5 INCH DISKETTE
? COMPUTER: IBM PS/2
? OPERATING SYSTEM: MS-DOS
? SOFTWARE: WORD PERFECT 5.1
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/462,169B
? FILING DATE: 05 JUN 95
? CLASSIFICATION: 514
? ATTORNEY/AGENT INFORMATION:
? NAME: MULLINS, J.G.
? REGISTRATION NUMBER: 33,073
? REFERENCE/DOCKET NUMBER: 325800-441 (PF203)
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 201-994-1700
? TELEFAX: 201-994-1744
? INFORMATION FOR SEQ ID NO: 5:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 35 BASE PAIRS
? TYPE: NUCLEIC ACID
? STRANDEDNESS: SINGLE
? TOPOLOGY: LINEAR
? MOLECULE TYPE: Oligonucleotide
? US-08-462-169B-5

Query Match 80.0% Score 14.47 DB 3 Length 35
Best Local Similarity 93.8% Pred. No. 1.8e+02
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0

QY : ATGCGGCTCCACTAG 16
DB : ATGCGGCTCCACTAG 1
|||||
16 ATGCGGCTCCACTAG 1

RESULT 3
US-08-462-169B-7/5
? Sequence 7, Application US/08462169B
? Patent No. 6773252
? GENERAL INFORMATION:
? APPLICANT: John Greene and Craig A. Roser
? TITLE OF INVENTION: Fibroblast Growth Factor
? NUMBER OF SEQUENCES: 32
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
? ADDRESSEE: CECCHI, STEWART & OLSTEIN
? STREET: 6 BECKER FARM ROAD
? CITY: ROSELAND
? STATE: NEW JERSEY
? COUNTRY: USA
? ZIP: 07068
? COMPUTER READABLE FORM:
? MEDIUM TYPE: 3.5 INCH DISKETTE
? COMPUTER: IBM PS/2
? OPERATING SYSTEM: MS-DOS
? SOFTWARE: WORD PERFECT 5.1
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/462,169B
? FILING DATE: 05 JUN 95
? CLASSIFICATION: 514
? ATTORNEY/AGENT INFORMATION:
? NAME: MULLINS, J.G.
? REGISTRATION NUMBER: 33,073
? REFERENCE/DOCKET NUMBER: 325800-441 (PF203)
? TELECOMMUNICATION INFORMATION:

```

```

? TELEPHONE: 201-994-1700
? TELEFAX: 201-994-1744
? INFORMATION FOR SEQ ID NO: 15:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 35 BASE PAIRS
? TYPE: NUCLEIC ACID
? STRANDEDNESS: SINGLE
? TOPOLOGY: LINEAR
? MOLECULE TYPE: (Oligonucleotide)
? US-08-462-169B-7

Query Match 80.0% Score 14.47 DB 3 Length 35
Best Local Similarity 93.8% Pred. No. 1.8e+02
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0

QY : ATGCGGCTCCACTAG 16
DB : ATGCGGCTCCACTAG 1
|||||
16 ATGCGGCTCCACTAG 1

RESULT 4
US-09-103-079-7/5
? Sequence 7, Application US/09103079A
? Patent No. 6334477
? GENERAL INFORMATION:
? APPLICANT: Greene, John M.
? TITLE OF INVENTION: Fibroblast Growth Factor
? FILE REFERENCE: PENDING
? CURRENT APPLICATION NUMBER: US/09/103,079A
? CURRENT FILING DATE: 1999-06-23
? EARLIER APPLICATION NUMBER: US/462,169
? EARLIER FILING DATE: 1995-06-25
? NUMBER OF SEQ ID NOS: 32
? SOFTWARE: Patent In Ver. 2.0
? SEQ ID NO: 7
? LENGTH: 35
? TYPE: DNA
? ORGANISM: Homo sapiens
? US-09-103-079-5

Query Match 80.0% Score 14.47 DB 3 Length 35
Best Local Similarity 93.8% Pred. No. 1.8e+02
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0

QY : ATGCGGCTCCACTAG 16
DB : ATGCGGCTCCACTAG 1
|||||
16 ATGCGGCTCCACTAG 1

RESULT 5
US-09-103-079-7/5
? Sequence 7, Application US/09103079A
? Patent No. 6334477
? GENERAL INFORMATION:
? APPLICANT: Greene, John M.
? TITLE OF INVENTION: Fibroblast Growth Factor
? FILE REFERENCE: PENDING
? CURRENT APPLICATION NUMBER: US/09/103,079A
? CURRENT FILING DATE: 1999-06-23
? EARLIER APPLICATION NUMBER: US/462,169
? EARLIER FILING DATE: 1995-06-25
? NUMBER OF SEQ ID NOS: 32
? SOFTWARE: Patent In Ver. 2.0
? SEQ ID NO: 7
? LENGTH: 35
? TYPE: DNA
? ORGANISM: Homo sapiens
? US-09-103-079-7

Query Match 80.0% Score 14.47 DB 3 Length 35
Best Local Similarity 93.8% Pred. No. 1.8e+02

```

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGCGGCTCCACTAG 16  
 DB 16 ATGGCGGCTCCACTAG 16

RESULT 6  
 US-09 425 021 5/0  
 : Sequence 5, Application US/09425001  
 : Patent No. 6482408  
 : GENERAL INFORMATION:  
 : APPLICANT: Greene, John V.  
 : APPLICANT: Rosen, Craig A.  
 : TITLE OF INVENTION: Fibroblast Growth Factor 15  
 : FILE REFERENCE: PF2001  
 : CURRENT APPLICATION NUMBER: US/09/425,021  
 : CURRENT FILING DATE: 1999-10-25  
 : EARLIER APPLICATION NUMBER: 09/101,070  
 : EARLIER FILING DATE: 1998-06-23  
 : NUMBER OF SEQ ID NOS: 32  
 : SOFTWARE: Patent In Ver. 2.0  
 : SEQ ID NO 5:  
 : LENGTH: 15  
 : TYPE: DNA  
 : ORGANISM: Homo sapiens  
 : US 09 425 021 5

Query Match 80.0%; Score 14.4; DB 4; Length 15;  
 Best Local Similarity 93.8%; Pred. No. 1.8e+02;  
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGCGGCTCCACTAG 16  
 DB 16 ATGGCGGCTCCACTAG 16

RESULT 7  
 US-09 425 021 5/0  
 : Sequence 7, Application US/09425001  
 : Patent No. 6482408  
 : GENERAL INFORMATION:  
 : APPLICANT: Greene, John V.  
 : APPLICANT: Rosen, Craig A.  
 : TITLE OF INVENTION: Fibroblast Growth Factor 15  
 : FILE REFERENCE: PF2001  
 : CURRENT APPLICATION NUMBER: US/09/425,021  
 : CURRENT FILING DATE: 1999-10-25  
 : EARLIER APPLICATION NUMBER: 09/101,070  
 : EARLIER FILING DATE: 1998-06-23  
 : NUMBER OF SEQ ID NOS: 32  
 : SOFTWARE: Patent In Ver. 2.0  
 : SEQ ID NO 7:  
 : LENGTH: 15  
 : TYPE: DNA  
 : ORGANISM: Homo sapiens  
 : US 09 425 021 7

Query Match 80.0%; Score 14.4; DB 4; Length 15;  
 Best Local Similarity 93.8%; Pred. No. 1.8e+02;  
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGCGGCTCCACTAG 16  
 DB 16 ATGGCGGCTCCACTAG 16

RESULT 8  
 US-09 006 932-10  
 : Sequence 10, Application US/09006932  
 : Patent No. 6239258  
 : GENERAL INFORMATION:  
 : APPLICANT: German, Michael

APPLICANT: Bell, Graeme I.  
 APPLICANT: Furuta, Hiroto  
 APPLICANT: Sussel, Lor:  
 TITLE OF INVENTION: Human NKx-2.2 Polypeptide-Encoding  
 Nucleotide Sequences  
 NUMBER OF SEQUENCES: 21  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Bozicovic & Reed, LLP  
 STREET: 285 Hamilton Ave, Suite 200  
 CITY: Palo Alto  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 94301  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSeq for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/008,892  
 FILING DATE:  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Francis, Carol L.  
 REGISTRATION NUMBER: 36,513  
 REFERENCE/DOCKET NUMBER: 9076-82CIP  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 650-327-3400  
 TELEFAX: 650-327-1231  
 TELEX:  
 INFORMATION FOR SEQ ID NO: 10:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 851 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-09 008 892-10

Query Match 72.8%; Score 14; DB 3; Length 851;  
 Best Local Similarity 100.0%; Pred. No. 3e+02;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CCGCTCGACTAGG 10  
 DB 10 CCGCTCGACTAGG 10

RESULT 9  
 US-09 060 756-49  
 : Sequence 49, Application US/09060756  
 : Patent No. 6183957  
 : GENERAL INFORMATION:  
 : APPLICANT: Cole, Stewart  
 : APPLICANT: Buchrieser Brosch, Roland  
 : APPLICANT: Gordon, Stephen  
 : APPLICANT: Billault, Alain  
 : TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM  
 THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA  
 TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA  
 FILE REFERENCE: 3495-0169  
 CURRENT APPLICATION NUMBER: US/09/060,756  
 CURRENT FILING DATE: 1998-04-16  
 NUMBER OF SEQ ID NOS: 743  
 SOFTWARE: Patent In Ver. 2.0  
 : SEQ ID NO 49:  
 : LENGTH: 461  
 : TYPE: DNA  
 : ORGANISM: Mycobacterium tuberculosis  
 : FEATURE:  
 : NAME/KEY: unseq

```

; LOCATION: (various positions within the sequence)
; OTHER INFORMATION: applicants are uncertain of bases designated as "n"
US-09-060-756 49

Query Match          76.7%; Score 13.8; DB 3; Length 461;
Best Local Similarity 88.2%; Pred. No. 3.7e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGCGGCTCCACTAGG 17
Db 325 ACGCGGCGCCACTAGG 34;

RESULT 10
US-09-670-314-49
; Sequence 49, Application US/09670314
; Patent No. 6492506
; GENERAL INFORMATION:
; APPLICANT: Cole, Stewart
; APPLICANT: Buchrieser Brosch, Roland
; APPLICANT: Gordon, Stephen
; APPLICANT: Billaut, Alain
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
; TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
; FILE REFERENCE: 3495-0169
; CURRENT APPLICATION NUMBER: US/09/670,314
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: 09/060,756
; PRIOR FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 49
; LENGTH: 461
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (various positions within the sequence)
; OTHER INFORMATION: applicants are uncertain of bases designated as "n"
US-09-670-314-49

Query Match          76.7%; Score 13.8; DB 4; Length 461;
Best Local Similarity 88.2%; Pred. No. 3.7e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGCGGCTCCACTAGG 17
Db 325 ACGCGGCGCCACTAGG 34;

RESULT 11
US-09-996-243-101/C
; Sequence 101, Application US/09996243
; Patent No. 6478825
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Bostein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: K-Javin, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas P.
```

```

; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Laurel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William L.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Segmented and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P730P1013
; CURRENT APPLICATION NUMBER: US/09/996,243
; CURRENT FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 82/043987
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 82/062240
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 86/065166
; PRIOR FILING DATE: 1993-11-10
; PRIOR APPLICATION NUMBER: 86/066911
; PRIOR FILING DATE: 1993-11-14
; PRIOR APPLICATION NUMBER: 86/066911
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: 87/075945
; PRIOR FILING DATE: 1993-02-25
; PRIOR APPLICATION NUMBER: 89/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/283322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 82/087156
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 82/087689
; PRIOR FILING DATE: 1998-06-07
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-07
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-07
; PRIOR APPLICATION NUMBER: 82/087927
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 62/098021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 80/089025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/288026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/288028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 62/088029
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 82/088030
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/288033
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 62/088426
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 82/088467
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 82/088467
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/288465
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/288734
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/288738
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/288742
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 62/088910
```

1 PRIOR FILING DATE: 1998-06-10  
2 PRIOR APPLICATION NUMBER: 60/088824  
3 PRIOR FILING DATE: 1998-06-10  
4 PRIOR APPLICATION NUMBER: 60/088826  
5 PRIOR FILING DATE: 1998-06-10  
6 PRIOR APPLICATION NUMBER: 60/088852  
7 PRIOR FILING DATE: 1998-06-11  
8 PRIOR APPLICATION NUMBER: 60/088861  
9 PRIOR FILING DATE: 1998-06-11  
10 PRIOR APPLICATION NUMBER: 60/088876  
11 PRIOR FILING DATE: 1998-06-11  
12 PRIOR APPLICATION NUMBER: 60/088905  
13 PRIOR FILING DATE: 1998-06-12  
14 PRIOR APPLICATION NUMBER: 60/089440  
15 PRIOR FILING DATE: 1998-06-16  
16 PRIOR APPLICATION NUMBER: 60/089511  
17 PRIOR FILING DATE: 1998-06-16  
18 PRIOR APPLICATION NUMBER: 60/089514  
19 PRIOR FILING DATE: 1998-06-16  
20 PRIOR APPLICATION NUMBER: 60/089532  
21 PRIOR FILING DATE: 1998-06-17  
22 PRIOR APPLICATION NUMBER: 60/089538  
23 PRIOR FILING DATE: 1998-06-17  
24 PRIOR APPLICATION NUMBER: 60/089568  
25 PRIOR FILING DATE: 1998-06-17  
26 PRIOR APPLICATION NUMBER: 60/089599  
27 PRIOR FILING DATE: 1998-06-17  
28 PRIOR APPLICATION NUMBER: 60/089606  
29 PRIOR FILING DATE: 1998-06-17  
30 PRIOR APPLICATION NUMBER: 60/089654  
31 PRIOR FILING DATE: 1998-06-17  
32 PRIOR APPLICATION NUMBER: 60/089661  
33 PRIOR FILING DATE: 1998-06-18  
34 PRIOR APPLICATION NUMBER: 60/089667  
35 PRIOR FILING DATE: 1998-06-18  
36 PRIOR APPLICATION NUMBER: 60/089698  
37 PRIOR FILING DATE: 1998-06-19  
38 PRIOR APPLICATION NUMBER: 60/089947  
39 PRIOR FILING DATE: 1998-06-19  
40 PRIOR APPLICATION NUMBER: 60/089948  
41 PRIOR FILING DATE: 1998-06-19  
42 PRIOR APPLICATION NUMBER: 60/089952  
43 PRIOR FILING DATE: 1998-06-19  
44 PRIOR APPLICATION NUMBER: 60/090246  
45 PRIOR FILING DATE: 1998-06-22  
46 PRIOR APPLICATION NUMBER: 60/090257  
47 PRIOR FILING DATE: 1998-06-22  
48 PRIOR APPLICATION NUMBER: 60/090354  
49 PRIOR FILING DATE: 1998-06-22  
50 PRIOR APPLICATION NUMBER: 60/090349  
51 PRIOR FILING DATE: 1998-06-23  
52 PRIOR APPLICATION NUMBER: 60/090355  
53 PRIOR FILING DATE: 1998-06-23  
54 PRIOR APPLICATION NUMBER: 60/090429  
55 PRIOR FILING DATE: 1998-06-24  
56 PRIOR APPLICATION NUMBER: 60/090431  
57 PRIOR FILING DATE: 1998-06-24  
58 PRIOR APPLICATION NUMBER: 60/090435  
59 PRIOR FILING DATE: 1998-06-24  
60 PRIOR APPLICATION NUMBER: 60/090444  
61 PRIOR FILING DATE: 1998-06-24  
62 PRIOR APPLICATION NUMBER: 60/090445  
63 PRIOR FILING DATE: 1998-06-24  
64 PRIOR APPLICATION NUMBER: 60/090472  
65 PRIOR FILING DATE: 1998-06-24  
66 PRIOR APPLICATION NUMBER: 60/090535  
67 PRIOR FILING DATE: 1998-06-24  
68 PRIOR APPLICATION NUMBER: 60/090540  
69 PRIOR FILING DATE: 1998-06-24  
70 PRIOR APPLICATION NUMBER: 60/090542  
71 PRIOR FILING DATE: 1998-06-24  
72 PRIOR APPLICATION NUMBER: 60/090557  
73 PRIOR FILING DATE: 1998-06-24

1 PRIOR APPLICATION NUMBER: 60/090676  
2 PRIOR FILING DATE: 1998-06-25  
3 PRIOR APPLICATION NUMBER: 60/090678  
4 PRIOR FILING DATE: 1998-06-25  
5 PRIOR APPLICATION NUMBER: 60/090690  
6 PRIOR FILING DATE: 1998-06-25  
7 PRIOR APPLICATION NUMBER: 60/090694  
8 PRIOR FILING DATE: 1998-06-25  
9 PRIOR APPLICATION NUMBER: 60/090655  
10 PRIOR FILING DATE: 1998-06-25  
11 PRIOR APPLICATION NUMBER: 60/090696  
12 PRIOR FILING DATE: 1998-06-25  
13 PRIOR APPLICATION NUMBER: 60/090862  
14 PRIOR FILING DATE: 1998-06-26  
15 PRIOR APPLICATION NUMBER: 60/090863  
16 PRIOR FILING DATE: 1998-06-26  
17 PRIOR APPLICATION NUMBER: 60/090863  
18 PRIOR FILING DATE: 1998-07-01  
19 PRIOR APPLICATION NUMBER: 60/091544  
20 PRIOR FILING DATE: 1998-07-01  
21 PRIOR APPLICATION NUMBER: 60/091519  
22 PRIOR FILING DATE: 1998-07-02  
23 PRIOR APPLICATION NUMBER: 60/091625  
24 PRIOR FILING DATE: 1998-07-02  
25 PRIOR APPLICATION NUMBER: 60/091633  
26 PRIOR FILING DATE: 1998-07-02  
27 PRIOR APPLICATION NUMBER: 60/091978  
28 PRIOR FILING DATE: 1998-07-02  
29 PRIOR APPLICATION NUMBER: 60/091982  
30 PRIOR FILING DATE: 1998-07-07  
31 PRIOR APPLICATION NUMBER: 60/092182  
32 PRIOR FILING DATE: 1998-07-09

Query Match 75.7%; Score 13.8; DA 4; Length 543;

Best Local Similarity 88.2%; Pred. No. 3.7e+02;

Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TGGGGGCTCCACTAGGG 18

DB 50 TGGGGGCTCCACTAGGG 34

RESULT 12

US-09-598-243--007c

Sequence 100, Application US/09096243

Patent No. 6478826

GENERAL INFORMATION:

APPLICANT: Ashkenazi, Avi J.

APPLICANT: Baker, Kevin P.

APPLICANT: Botstein, David

APPLICANT: Cessoyers, Sue

APPLICANT: Eaton, Dan L.

APPLICANT: Ferrara, Napoleone

APPLICANT: Fogel, Sherman

APPLICANT: Gerber, Hanspeter

APPLICANT: Gerlicsen, Mary E.

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Grimaldi, J. Christopher

APPLICANT: Gurley, Austin L.

APPLICANT: Kijavitsky, J.

APPLICANT: Napier, Mary A.

APPLICANT: Pan, James

APPLICANT: Pacht, Nicholas P.

APPLICANT: Roy, Margaret Ann

APPLICANT: Stewart, Timothy A.

APPLICANT: Tomas, Daniel

APPLICANT: Watanabe, Colin K.

APPLICANT: Williams, P. Mickey

APPLICANT: Wood, William J.

APPLICANT: Zhang, Zemin

1 TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleo  
2  
3 TITLE OF INVENTION: Acids Encoding the Same  
4  
5 FILE REFERENCE: P2730PIC13  
6  
7 CURRENT APPLICATION NUMBER: US/09/996,243  
8  
9 CURRENT FILING DATE: 2001-11-14  
10  
11 PRIOR APPLICATION NUMBER: 60/349787  
12  
13 PRIOR FILING DATE: 1997-06-16  
14  
15 PRIOR APPLICATION NUMBER: 60/062250  
16  
17 PRIOR FILING DATE: 1997-10-17  
18  
19 PRIOR APPLICATION NUMBER: 60/065186  
20  
21 PRIOR FILING DATE: 1997-11-12  
22  
23 PRIOR APPLICATION NUMBER: 60/365311  
24  
25 PRIOR FILING DATE: 1997-11-13  
26  
27 PRIOR APPLICATION NUMBER: 60/066770  
28  
29 PRIOR FILING DATE: 1997-11-24  
30  
31 PRIOR APPLICATION NUMBER: 60/075945  
32  
33 PRIOR FILING DATE: 1998-02-25  
34  
35 PRIOR APPLICATION NUMBER: 60/078910  
36  
37 PRIOR FILING DATE: 1998-03-20  
38  
39 PRIOR APPLICATION NUMBER: 60/083322  
40  
41 PRIOR FILING DATE: 1998-04-28  
42  
43 PRIOR APPLICATION NUMBER: 60/084600  
44  
45 PRIOR FILING DATE: 1998-05-07  
46  
47 PRIOR APPLICATION NUMBER: 60/087106  
48  
49 PRIOR FILING DATE: 1998-05-28  
50  
51 PRIOR APPLICATION NUMBER: 60/087607  
52  
53 PRIOR FILING DATE: 1998-06-02  
54  
55 PRIOR APPLICATION NUMBER: 60/087609  
56  
57 PRIOR FILING DATE: 1998-06-02  
58  
59 PRIOR APPLICATION NUMBER: 60/087759  
60  
61 PRIOR FILING DATE: 1998-06-02  
62  
63 PRIOR APPLICATION NUMBER: 60/087827  
64  
65 PRIOR FILING DATE: 1998-06-03  
66  
67 PRIOR APPLICATION NUMBER: 60/288021  
68  
69 PRIOR FILING DATE: 1998-06-04  
70  
71 PRIOR APPLICATION NUMBER: 60/088325  
72  
73 PRIOR FILING DATE: 1998-06-04  
74  
75 PRIOR APPLICATION NUMBER: 60/088026  
76  
77 PRIOR FILING DATE: 1998-06-04  
78  
79 PRIOR APPLICATION NUMBER: 60/068028  
80  
81 PRIOR FILING DATE: 1998-06-04  
82  
83 PRIOR APPLICATION NUMBER: 60/288029  
84  
85 PRIOR FILING DATE: 1998-06-04  
86  
87 PRIOR APPLICATION NUMBER: 60/088030  
88  
89 PRIOR FILING DATE: 1998-06-04  
90  
91 PRIOR APPLICATION NUMBER: 60/088033  
92  
93 PRIOR FILING DATE: 1998-06-04  
94  
95 PRIOR APPLICATION NUMBER: 60/088326  
96  
97 PRIOR FILING DATE: 1998-06-04  
98  
99 PRIOR APPLICATION NUMBER: 60/088167  
100  
101 PRIOR FILING DATE: 1998-06-05  
102  
103 PRIOR APPLICATION NUMBER: 60/088202  
104  
105 PRIOR FILING DATE: 1998-06-05  
106  
107 PRIOR APPLICATION NUMBER: 60/088212  
108  
109 PRIOR FILING DATE: 1998-06-05  
110  
111 PRIOR APPLICATION NUMBER: 60/088217  
112  
113 PRIOR FILING DATE: 1998-06-05  
114  
115 PRIOR APPLICATION NUMBER: 60/088655  
116  
117 PRIOR FILING DATE: 1998-06-09  
118  
119 PRIOR APPLICATION NUMBER: 60/088714  
120  
121 PRIOR FILING DATE: 1998-06-10  
122  
123 PRIOR APPLICATION NUMBER: 60/088718  
124  
125 PRIOR FILING DATE: 1998-06-10  
126  
127 PRIOR APPLICATION NUMBER: 60/088742  
128  
129 PRIOR FILING DATE: 1998-06-10  
130  
131 PRIOR APPLICATION NUMBER: 60/088810  
132  
133 PRIOR FILING DATE: 1998-06-10  
134  
135 PRIOR APPLICATION NUMBER: 60/088824  
136  
137 PRIOR FILING DATE: 1998-06-10  
138  
139 PRIOR APPLICATION NUMBER: 60/088826  
140  
141 PRIOR FILING DATE: 1998-06-10  
142  
143 PRIOR APPLICATION NUMBER: 60/088858  
144  
145 PRIOR FILING DATE: 1998-06-11  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000







US-08 576-626A-1

Query Match 76.7% Score 13.87 BR 27 Length 8746  
 Best Local Similarity 89.2% Pred. No. 3.4e+02  
 Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0

QY 2 TGGCGGCTCCACTAGG 16  
 |||||  
 DT 534 TGGCGGCTCCACTAGG 512

RESULT 15

US-08 576-626A-1  
 ? Sequence 1: Application US/9576-26A  
 ? Patent No. 5698194  
 ? GENERAL INFORMATION:  
 ? APPLICANT: SUMMERS, R.G.  
 ? APPLICANT: KATZ, L.  
 ? APPLICANT: DONADIO, S.  
 ? APPLICANT: STAYER, M.C.  
 ? TITLE OF INVENTION: POLYKETIDE-ASSOCIATED SUGAR  
 ? TITLE OF INVENTION: BIOSYNTHESIS GENES  
 ? NUMBER OF SEQUENCES: 60  
 ? CORRESPONDENCE ADDRESS:  
 ? ADDRESSEE: Abbott Laboratories  
 ? STREET: 100 Abbott Park Road  
 ? CITY: Abbott Park  
 ? STATE: Illinois  
 ? COUNTRY: USA  
 ? ZIP: 60064-3502  
 ? COMPUTER READABLE FORM:  
 ? MEDIUM TYPE: Diskette  
 ? COMPUTER: IBM Compatible  
 ? OPERATING SYSTEM: DOS  
 ? SOFTWARE: FASTSEQ Version 2.0  
 ? CURRENT APPLICATION DATA:  
 ? APPLICATION NUMBER: US/9576-626A  
 ? FILING DATE: 21-DEC-1995  
 ? CLASSIFICATION: 435  
 ? PRIOR APPLICATION DATA:  
 ? APPLICATION NUMBER:  
 ? FILING DATE:  
 ? ATTORNEY/AGENT INFORMATION:  
 ? NAME: Dianne Casuto  
 ? REGISTRATION NUMBER: 140,043  
 ? REFERENCE/SOCKET NUMBER: 569,194-1  
 ? TELECOMMUNICATION INFORMATION:  
 ? TELEPHONE: (847) 938-2157  
 ? TELEFAX: (847) 938-2223  
 ? TELEX:  
 ? INFORMATION FOR SEQ ID NO: 1:  
 ? SEQUENCE CHARACTERISTICS:  
 ? LENGTH: 3756 base pairs  
 ? TYPE: nucleic acid  
 ? STRANDEDNESS: single  
 ? TOPOLOGY: linear  
 US-08 576-626A-1

Query Match 76.7% Score 13.87 BR 27 Length 8746  
 Best Local Similarity 89.2% Pred. No. 3.4e+02  
 Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0

QY 2 TGGCGGCTCCACTAGG 16  
 |||||  
 DT 534 TGGCGGCTCCACTAGG 512

Search completed: October 21, 2003, 05:19:43  
 Job time: 8.414s secs

GenCore version: 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

CM nucleic - nucleic search, using sw model

Run on: October 20, 2003, 17:12:14 : Search time 10.1934 Seconds  
without alignments:  
4672.369 Million cell updates/sec

Target: US-10-029-137-8  
Perfect score: 18  
Sequence: 1 atggcgctcactagg33 18

Scoring table: IDENTITY NUC  
Gapco 10.0 : Gapco 1.0

Searches: 252756 seqs, 1349713017 residues  
Total number of hits satisfying chosen parameters: 5105512

Minimum hit seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :	N Geneset	Accession
1:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:	
2:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:	
3:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:	
4:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:	
5:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:	
6:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:	
7:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:	
8:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:	
9:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:	
10:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:	
11:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:	
12:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:	
13:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:	
14:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:	
15:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:	
16:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:	
17:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:	
18:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:	
19:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:	
20:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:	
21:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:	
22:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:	
23:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:	
24:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:	
25:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:	

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY:ES

Result No.	Score	Query Match	Length	ID	Description
1	18	100.0	465	21	Human secreted pro
2	18	100.0	862	20	Human ovarian tumo
3	18	100.0	1071	20	Human breast tumo
4	18	100.0	10995	25	Secreted protein 9
5	16.4	91.1	549	21	Arabidopsis thalia
6	16.4	91.1	1737	21	Arabidopsis thalia
7	16.4	91.1	2440	21	Arabidopsis thalia
8	16	89.9	244	22	Probe seq. Synthe

9	15.4	55.6	1837	23	ABI26433
10	15.4	85.6	4024	23	ABL26432
11	14.8	82.2	77	22	ABAS1460
12	14.8	82.2	77	22	ABAS1461
13	14.8	82.2	323	21	AAZ16683
14	14.8	82.2	446	22	ABA46358
15	14.8	82.2	446	22	ABA26536
16	14.8	82.2	463	21	AAZ16705
17	14.8	82.2	855	20	AAV80618
18	14.8	82.2	940	24	ASK63794
19	14.8	82.2	10028	22	AAK68407
20	14.8	82.2	13795	22	AAK06934
21	14.8	82.2	13795	22	AAK1474
22	14.8	82.2	13795	22	AAK33116
23	14.8	82.2	13795	24	APV94172
24	14.8	82.2	13795	24	ABQ66798
25	14.8	82.2	249487	24	ABN85733
26	14.4	80.0	35	18	AAT47726
27	14.4	80.0	35	19	AAV36771
28	14.4	80.0	35	19	AAV36773
29	14.4	80.0	35	21	AAK02226
30	14.4	80.0	35	21	AAK02228
31	14.4	80.0	35	21	AAK28156
32	14.4	80.0	35	21	AAK28158
33	14.4	80.0	35	24	ABQ78009
34	14.4	80.0	332	24	ABQ78162
35	14.4	80.0	412	22	ABA76709
36	14.4	80.0	412	22	ABA41207
37	14.4	80.0	412	22	AAK25337
38	14.4	80.0	412	22	AAK51345
39	14.4	80.0	412	22	AAI28360
40	14.4	80.0	412	22	AAI57414
41	14.4	80.0	412	23	ASB50893
42	14.4	80.0	412	24	ASB24891
43	14.4	80.0	578	22	ASB53525
44	14.4	80.0	578	22	ABA30721
45	14.4	80.0	578	22	AAK12052

ALIGNMENTS

RESULT 1

AAK01734  
ID AAC01734 standard; cDNA; 465 bp.

AC AAC01734;

CT CG CAT-2600 (first entry);

DE Human secreted protein 5' EST, SEQ ID NO: 1792.

Human; 5' EST: expressed sequence tag; secreted protein; cDNA isolation;  
gene therapy; chromosome mapping; ss.

CS Homo sapiens.

PN EP1033401-A2.

PD 36-SEP-2000.

PF 21-FEB-2000, 2000EP-0205610.

PR 26-FEB-1999; 99US-0122487.

XX (GEST ) GENSET.

PI Damas Malne Edwards J, Duciet A, Giordano C;

DR WPI: 2000-500351/45.

XX P-PSDB; AAG01769.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST: for

PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 PS Claim 1: SEQ ID 192; 71bp + CD-ROM; English.  
 XX The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. An ORF has been identified within the  
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
 CC derived from 30 different tissues. EST sequences usually correspond  
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
 CC well suited for isolating cDNA sequences derived from the 5' ends of  
 CC mRNAs and even in those cases where longer cDNA sequences have been  
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length  
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
 CC gene therapy and chromosome mapping procedures. They are used to obtain  
 CC upstream regulatory sequences and to design expression and secretion  
 CC vectors.  
 XX Sequence 46: 82; 116 A; 113 G; 122 G; 104 T; 6 other;

Query Match 100.0%; Score 18; DB 20; Length 45;

Best Local Similarity 100.0%; Pred. No. 9.6; Mismatches 0; Indels 0; Gaps 0;

1 ATGGCGCTCCACTAGG 18

20 ATGGCGCTCCACTAGG 37

#### RESULT 2

ID AA277452 standard; cDNA; 882 BP.

AA277452;

10 APR-2000 (first entry)

Human ovarian tumor cDNA library derived EST fragment 3.

Expressed sequence tag; EST; human; ovarian tumor; anticancer;

gene therapy; treatment; ss.

Homo sapiens.

DE19817557 AL.

21-OCT-1999.

09-APR-1996; 98DE-1017557.

09-APR-1998; 98DE 1017557.

(META) METAGEN GES GENOMFORSCHUNG MBH.

Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;

WPI; 1999-531920/51.

P-PSDB; AA775573.

New nuclear acid sequences expressed in ovarian, and some other, cancer  
 tissues, and derived polypeptides, for treatment of ovarian cancer and  
 identification of therapeutic agents.

Claim 3; Page 140; 3:0pp; German.

This invention describes novel nucleic acid (cDNA) sequences (A) which  
 have anticancer activity and are highly expressed in ovarian tumor  
 tissue (and some also in testis and breast cancer tissue). The products  
 of the invention can be used for gene therapy. (A) are used (i) for  
 recombinant expression of polypeptides (B) and (ii) to isolate complete  
 genes. (B), are used (i) to identify agents suitable for treatment of

ovarian cancer; (ii) directly for treating this form of cancer  
 including expression from gene therapy vectors) and (iii) for generation  
 of specific antibodies. (A) are identified by assembling ESTs (expressed  
 sequence tags) from a particular tissue type before comparison of the  
 expression patterns. This allows a significantly longer fragment of the  
 gene to be revealed, so should reduce the number of failures associated  
 with the fact that ESTs from different libraries may represent different  
 parts of the same known gene, distorting the estimated frequency of  
 occurrence in a particular tissue. AA277450-277572 represent the human  
 ovarian tumor cDNA library derived EST fragments described in the method  
 of the invention and encode the protein fragments represented in  
 AA277505-277636.

Sequence 882 BP; 218 A; 190 G; 104 G; 248 T; 0 other;

Query Match 100.0%; Score 18; DB 20; Length 882;

Best Local Similarity 100.0%; Pred. No. 10;

Mismatches 0; Indels 0; Gaps 0;

1 ATGGCGCTCCACTAGG 18

20 ATGGCGCTCCACTAGG 37

#### RESULT 3

ID AA233600 standard; cDNA; 1071 BP.

AA233600;

08-DEC-1999 (first entry)

Human breast tumour-associated EST 60.

Expressed sequence tag; EST; human; breast cancer; cytostatic;  
 medicaments; gene therapy; treatment; fat metabolism; ss.

Homo sapiens.

DE19813835 AL.

21-SEP-1999.

20-MAR-1999; 98DE 1017557.

20-MAR-1999; 98DE 1017557.

(META) METAGEN GES GENOMFORSCHUNG MBH.

Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;

WPI; 1999-529979/45.

P-PSDB; AA233606.

Human nuclear acid sequences and protein products from normal breast  
 tissue, useful for breast cancer therapy

Claim 1a; 145; 200pp; German.

This invention describes novel human nucleic acid sequences from normal  
 breast tissue which have tyrosinase activity. The nucleic acid sequences  
 can be used to produce and isolate full-length gene sequences. They can  
 be used to express proteins, which can be used as tools to find an  
 activity against breast cancer. The sequences can be used in sense or  
 antisense form. They are especially useful for medicaments for gene  
 therapy to treat breast cancer and for treating illnesses associated  
 with fat metabolism. AA233591-233610 represent expressed sequence tags  
 described in the method of the invention.

Sequence 1071 BP; 281 A; 218 G; 255 G; 317 T; 0 other;

Query Match 100.0%; Score 18; DB 20; Length 1071;

Best Local Similarity 100.0%; Pred. No. 10;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY 1 ATGGCGGTCCTCACTAGG 18  
||||| ||||| |||  
DB 10 ATGGCGGTCCTCACTAGG 27

RESULT 4  
ABZ74584  
10 ABZ74584 standard; DNA; 10995 BP.  
XX  
AC ABZ74584;  
XX  
DT 10 MAY 2003 (first entry)  
XX  
DE Secreted protein gene 158 genomic fragment HX0736, SEQ ID NO:1731.  
XX  
KW Human; secreted protein; cancer; tumor; hyperproliferative disorder;  
KW autoimmune disorder; inflammation; angiogenic diseases; AIDS;  
KW acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;  
KW drug screening; chromosome identification; chromosome mapping;  
KW cytostatic; gene therapy; antiinflammatory; immunomodulator; anti-HIV;  
KW antitumoric; vulnery; chromosome 4p16; gene; ds.  
XX  
OS Homo sapiens.  
XX  
FN W010277013-A2.  
XX  
PD 01-JUN 2002.  
XX  
PF 26 MAR 2002; 2002WC-US0370.  
XX  
PE 27 MAR 2001; 2001US-278650P.  
PR 12 SEP 2001; 2001US-0950592.  
PR 12-SEP 2001; 2001US-0950593.  
XX  
RA HUMA : HUMAN GENOME SCI INC.  
XX  
RT Rosen CA, Rubin SX;  
XX  
DB WPI: 2001-040578/03.  
XX  
PT New human secreted proteins and nucleic acids, useful for detecting or  
PT treating cancer or other hyperproliferative disorders, autoimmune  
PT disorders, inflammatory disorders, HIV disease, hepatitis or anemia -  
XX Disclosure, Page 211, col. 14-15, figs 1-10.  
XX  
CC ABZ74581-ABZ74597 represent cDNAs of the invention. 141 human secreted  
CC protein genes, and ABZ00347-ABZ00349 represent the proteins they encode.  
CC ABZ74598-ABZ74657 represent human secreted protein genomic fragments. The  
CC invention also encompasses antibodies specific for the secreted proteins,  
CC the use of the secreted proteins in drug screening and recombinant  
CC vectors and host cells comprising a nucleic acid of the invention. The  
CC secreted proteins are thought to be involved in biological activities  
CC associated with cellular signalling, cellular differentiation, cell  
CC migration, prohormone activation and neurotransmitter activity. The  
CC secreted proteins, nucleic acids encoding them, antibodies or antibody  
CC fragments specific for the secreted proteins, and modulators of protein  
CC activity are useful for diagnosing or treating cancers or other  
CC hyperproliferative disorders. Additionally, the secreted proteins and  
CC their nucleic acids may also be used in the treatment of autoimmune  
CC disorders, inflammatory disorders, diseases involving angiogenesis, AIDS  
CC acquired immunodeficiency syndrome, hepatitis, anaemia, and to promote  
CC wound healing. Nucleic acids of the invention may be used for chromosome  
CC identification, chromosome mapping, in gene therapy, for identifying  
CC individuals from minute biological samples, as hybridisation probes, and  
CC as molecular weight markers. The present sequence represents a human  
CC secreted protein genomic fragment referred to in the disclosure of the  
XX invention.  
XX  
SQ Sequence 10995 BP; 2602 A; 2171 C; 2559 G; 1049 T; 0 other;

Query Match 100.0%; Score 18; DB 25; Length 10995;  
Best Local Similarity 100.0%; Prod. No. 12;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCGGTCCTCACTAGG 18  
||||| ||||| |||  
DB 10 ATGGCGGTCCTCACTAGG 12

RESULT 5  
AAC43655  
10 AAC43655 standard; DNA; 549 BP.  
XX  
AC AAC43655;  
XX  
DT 18 OCT 2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 40014.  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
FN EP1013445 A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PE 23-FEB 1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0121180.  
PR 09-MAR-1999; 99US-0123548.  
PR 21-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 14-APR-1999; 99US-0128845.  
PR 19-APR-1999; 99US-0130072.  
PR 23-APR-1999; 99US-0130449.  
PR 25-APR-1999; 99US-0130510.  
PR 29-APR-1999; 99US-0130892.  
PR 30-APR-1999; 99US-0131449.  
PR 31-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132497.  
PR 04-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132493.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134258.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135153.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137522.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138034.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN 1999; 99US-0138847.



```

RESULT: C
AAC43019
ID AAC43019 standard; DNA, 1737 BP.
XX
AC AAC43019;
XX
ET 17 OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 31861.
XX
KW Hybridisation assay; Genetic mapping; Gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
XX EPIC13406 A2.
XX
XX 24 SEP 2000.
XX
XX 25 FEB-2000; 2000EP-3301319.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 21-MAR-1999; 99US-0125388.
PR 25-MAR-1999; 99US-0126264.
PR 25-MAR-1999; 99US-0126785.
PR 31-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128214.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130377.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132488.
PR 12-MAY-1999; 99US-0133456.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134368.
PR 19-MAY-1999; 99US-0134441.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137322.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139464.
PR 18-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139767.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140981.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142255.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143442.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144326.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 23-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.

```

```

PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149729.
PR 23-AUG-1999; 99US-0149952.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150584.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151301.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155653.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156536.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158923.
PR 13-OCT-1999; 99US-0159234.
PR 13-OCT-1999; 99US-0159235.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159984.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161434.
PR 25-OCT-1999; 99US-0161435.
PR 25-OCT-1999; 99US-0161436.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161923.
PR 28-OCT-1999; 99US-0161932.
PR 28-OCT-1999; 99US-0161933.
PR 29-OCT-1999; 99US-0162142.

```

Query March 91.13; Score 16.4; EB 21; Length 1737;

Best Local Similarity 94.4;

Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 ATGGCGGCTCCACTAGG 18

14 ATGGCGGCTCCACTAGG 3;

14 ATGGCGGCTCCACTAGG 3;

RESULT 7

AAC51802

1D AAC51802 standard; DNA; 2440 BP.

XX

AC AAC51802;

```

XX 18-OCT-2000 (first entry)
DT Arabidopsis thaliana DNA fragment SEQ ID NO: 65782.
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 65782.
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX EP1033405 A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 200CEP-0139419.
XX 25-FEB-1999; 99US-0139425.
XX 25-FEB-1999; 99US-0139426.
XX 25-FEB-1999; 99US-0139427.
XX 25-FEB-1999; 99US-0139428.
XX 25-FEB-1999; 99US-0139429.
XX 25-FEB-1999; 99US-0139430.
XX 25-FEB-1999; 99US-0139431.
XX 25-FEB-1999; 99US-0139432.
XX 25-FEB-1999; 99US-0139433.
XX 25-FEB-1999; 99US-0139434.
XX 25-FEB-1999; 99US-0139435.
XX 25-FEB-1999; 99US-0139436.
XX 25-FEB-1999; 99US-0139437.
XX 25-FEB-1999; 99US-0139438.
XX 25-FEB-1999; 99US-0139439.
XX 25-FEB-1999; 99US-0139440.
XX 25-FEB-1999; 99US-0139441.
XX 25-FEB-1999; 99US-0139442.
XX 25-FEB-1999; 99US-0139443.
XX 25-FEB-1999; 99US-0139444.
XX 25-FEB-1999; 99US-0139445.
XX 25-FEB-1999; 99US-0139446.
XX 25-FEB-1999; 99US-0139447.
XX 25-FEB-1999; 99US-0139448.
XX 25-FEB-1999; 99US-0139449.
XX 25-FEB-1999; 99US-0139450.
XX 25-FEB-1999; 99US-0139451.
XX 25-FEB-1999; 99US-0139452.
XX 25-FEB-1999; 99US-0139453.
XX 25-FEB-1999; 99US-0139454.
XX 25-FEB-1999; 99US-0139455.
XX 25-FEB-1999; 99US-0139456.
XX 25-FEB-1999; 99US-0139457.
XX 25-FEB-1999; 99US-0139458.
XX 25-FEB-1999; 99US-0139459.
XX 25-FEB-1999; 99US-0139460.
XX 25-FEB-1999; 99US-0139461.
XX 25-FEB-1999; 99US-0139462.
XX 25-FEB-1999; 99US-0139463.

```





OS Synthetic.  
 XX JP2001245590-A.  
 XX  
 XX 11-SEP-2001.  
 XX  
 XX 03 MAR-2000; 2000JP-0059239.  
 XX  
 XX 03-MAR-2000; 2000JP-0059239.  
 XX  
 XX (NORQ) NORINSUISANSO SAKSHI KONCHU.  
 XX  
 XX WPI; 2001-K10C71/70.  
 XX  
 XX A silkworm grade for producing sericin in a large amount  
 XX  
 XX Example 3; Page 12; 17pp; Japanese.  
 XX  
 XX The present invention relates to a silkworm grade for producing sericin  
 XX in a large amount. The silkworm is obtained by crossing a normal grade of  
 XX strong high silk quantity line to a mutant grade of naked chrysalis line.  
 XX The present sequence is a probe, which was used in an example from the  
 XX present invention.  
 XX  
 XX Sequence 244 BP; (1 A; 70 C; 47 G; 46 T; 0 other).  
 XX  
 XX  
 XX Query Match 86.9%; Score 167; DB 24; Length 244;  
 XX Best Local Similarity 100.0%; Pred. No. 1e-02;  
 XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 XX : ATGGCGGCTCCACTAG 16  
 XX ||||| ||||| |||||  
 XX 184 ATGGCGGCTCCACTAG 169  
 XX  
 XX RESULT 9  
 XX ABL26433  
 XX ID ABL26433 standard; DNA; 1837 BP.  
 XX  
 XX ABL26433;  
 XX  
 XX 26-MAR-2003 (first entry);  
 XX  
 XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 10772.  
 XX  
 XX Drosophila; developmental biology; cell signalling; insecticide;  
 XX pharmacological; gene; ds.  
 XX Drosophila melanogaster.  
 XX  
 XX WC200101047 A2.  
 XX  
 XX 27-SEP-2001.  
 XX  
 XX 23-MAR-2001; 2001WO-US03231.  
 XX  
 XX 23 MAR-2000; 2000US 191637P.  
 XX  
 XX 11-JUL-2000; 2000US-C62415C.  
 XX  
 XX (PEKE) PE CORP NY.  
 XX  
 XX Venter JC, Adams M, Li PKL, Myers EW;  
 XX  
 XX WPI; 2001-K56A60/75.  
 XX  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 XX genes from Drosophila and for elucidating cell signalling and cell cell  
 XX interactions.  
 XX  
 XX Claim 1; SEQ ID NO 10772; 21pp + Sequence Listing; English.  
 XX  
 XX The invention relates to an isolated nucleic acid detection reagent  
 XX capable of detecting 1000 or more genes from Drosophila. The invention is  
 XX useful in developmental biology and in elucidating cell signalling and  
 XX cell-cell interactions in higher eukaryotes for the development of  
 XX insecticides, therapeutics and pharmaceutical drugs. The invention  
 XX discloses genomic DNA sequences (AB10176-AB10511), expressed DNA  
 XX sequences (AB101840-AB101875), and the encoded proteins  
 XX (AB101737-AB102072).  
 XX  
 XX The sequence data for this patent did not form part of the printed  
 XX specification, but was obtained in electronic format directly from WHO  
 XX at ftp:who.int/pub/publistmed/pub\_sequences.  
 XX  
 XX Sequence 4024 BP; 1041 A; 970 C; 984 G; 1019 T; 0 other;  
 XX  
 XX Query Match 85.8%; Score 154; DB 23; Length 4024;  
 XX Best Local Similarity 94.1%; Pred. No. 2.4e-02;

CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB10176-AB10511), expressed DNA  
 CC sequences (AB101840-AB101875), and the encoded proteins  
 CC (AB101737-AB102072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WHO  
 CC at ftp:who.int/pub/publistmed/pub\_sequences.  
 XX  
 XX Sequence 1837 BP; 478 A; 485 C; 477 G; 356 T; 0 other;  
 XX  
 XX Query Match 85.8%; Score 154; DB 23; Length 1837;  
 XX Best Local Similarity 94.1%; Pred. No. 2.4e-02;  
 XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 XX : ATGGCGGCTCCACTAG 17  
 XX ||||| ||||| |||||  
 XX 1557 ATGGCGGCTCCACTAG 1574  
 XX  
 XX RESULT 10  
 XX ABL26432  
 XX ID ABL26432 standard; DNA; 1804 BP.  
 XX  
 XX ABL26432;  
 XX  
 XX 26-MAR-2003 (first entry);  
 XX  
 XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 10774.  
 XX  
 XX Drosophila; developmental biology; cell signalling; insecticide;  
 XX pharmacological; gene; ds.  
 XX Drosophila melanogaster.  
 XX  
 XX WC200101042 A2.  
 XX  
 XX 27-SEP-2001.  
 XX  
 XX 23-MAR-2001; 2001WO-US03231.  
 XX  
 XX 23-MAR-2000; 2000US 191637P.  
 XX  
 XX 11-JUL-2000; 2000US-C62415C.  
 XX  
 XX (PEKE) PE CORP NY.  
 XX  
 XX Venter JC, Adams M, Li PKL, Myers EW;  
 XX  
 XX WPI; 2001-K56A60/75.  
 XX  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 XX genes from Drosophila and for elucidating cell signalling and cell cell  
 XX interactions.  
 XX  
 XX Claim 1; SEQ ID NO 10774; 21pp + Sequence Listing; English.  
 XX  
 XX The invention relates to an isolated nucleic acid detection reagent  
 XX capable of detecting 1000 or more genes from Drosophila. The invention is  
 XX useful in developmental biology and in elucidating cell signalling and  
 XX cell-cell interactions in higher eukaryotes for the development of  
 XX insecticides, therapeutics and pharmaceutical drugs. The invention  
 XX discloses genomic DNA sequences (AB10176-AB10511), expressed DNA  
 XX sequences (AB101840-AB101875), and the encoded proteins  
 XX (AB101737-AB102072).  
 XX  
 XX The sequence data for this patent did not form part of the printed  
 XX specification, but was obtained in electronic format directly from WHO  
 XX at ftp:who.int/pub/publistmed/pub\_sequences.  
 XX  
 XX Sequence 4024 BP; 1041 A; 970 C; 984 G; 1019 T; 0 other;  
 XX  
 XX Query Match 85.8%; Score 154; DB 23; Length 4024;  
 XX Best Local Similarity 94.1%; Pred. No. 2.4e-02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCGGCTCCACTAGG 18  
 D6 1 ATGGCGGCTCCACTAGG 18  
 ||||| ||||| |||||

RESULT 12  
 ABA36411  
 ID ABA36411 standard; cDNA; 77 BP.  
 XX  
 AC ABA36411;  
 XX  
 DT 23-JAN-2002 (first entry)  
 XX  
 DE Probe #14977 for gene expression analysis in human heart cell sample.  
 XX  
 KW Human; gene expression; heart; microarray; vascular system; probe;  
 KW cardiovascular disease; hypertension; cardiac arrhythmia;  
 KW congenital heart disease; ss.  
 XX  
 QS Homo sapiens.  
 XX  
 PN WC2002157274-A2.  
 XX  
 PD 09-AUG 2001.  
 XX  
 PF 30-JAN-2001; 2001WO 0500466.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 01-AUG 2000; 2000US-0632366.  
 PR 21-SEP 2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0234659.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLF 1) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel BK, Chen W, Rank DR;  
 XX  
 DP WPI; 2001-498899/53.  
 XX  
 PT New spatially addressable set of single exon nucleic acid probes,  
 PT useful for measuring gene expression in sample derived from human  
 PT breast. Contains number of single exon nucleic acid probes  
 XX  
 PS Claim 4; SEQ ID NO 101-5; 3200; a sequence derived from human  
 XX  
 SS The invention relates to a spatially addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human breast and B1 474 cells. The method involves contacting  
 CC the probes with a collection of detectably labeled nucleic acids  
 CC derived from mRNA of human breast, and then measuring the label  
 CC bound to each probe of the microarray. The probes are useful for  
 CC verifying the expression of regions of genomic DNA predicted to  
 CC encode proteins, they are useful for gene discovery, and for  
 CC determining predisposition and/or prognostic breast disease. Gene  
 CC expression analysis is useful for assessing the toxicity of chemical  
 CC agents on cells. The microarray of this invention presents a far greater  
 CC diversity of probes for measuring gene expression with far less bias  
 CC than expressed sequence tag microarrays. The method is suitable for  
 CC rapid production of functional information from genomic sequences. The  
 CC present sequence is a single exon nucleic acid probe of the invention.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 77 BP; 8 A; 15 C; 17 G; 37 T; 0 other;

Query Match 82.2%; Score 14.5; DB 22; Length 77;  
 Best Local Similarity 88.9%; Pred. No. 4e-02;  
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCGGCTCCACTAGG 18  
 D6 1 ATGGCGGCTCCACTAGG 18  
 ||||| ||||| |||||

RESULT 13  
 AAC16683  
 ID AAC16683 standard; cDNA; 321 BP.

QY 1 ATGGCGGCTCCACTAGG 18  
 D6 1 ATGGCGGCTCCACTAGG 18  
 ||||| ||||| |||||

RESULT 12  
 ABA36411  
 ID ABA36411 standard; cDNA; 77 BP.  
 XX  
 AC ABA36411;  
 XX  
 DT 23-JAN-2002 (first entry)  
 XX  
 DE Probe #14977 for gene expression analysis in human heart cell sample.  
 XX  
 KW Human; gene expression; heart; microarray; vascular system; probe;  
 KW cardiovascular disease; hypertension; cardiac arrhythmia;  
 KW congenital heart disease; ss.  
 XX  
 QS Homo sapiens.  
 XX  
 PN WC2002157274-A2.  
 XX  
 PD 09-AUG 2001.  
 XX  
 PF 30-JAN-2001; 2001WO 0500466.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 01-AUG 2000; 2000US-0632366.  
 PR 21-SEP 2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0234659.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLF 1) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel BK, Chen W, Rank DR;  
 XX  
 DP WPI; 2001-498899/53.  
 XX  
 PT Single exon nucleic acid probes for analyzing gene expression in human  
 PT hearts.  
 XX  
 PS Claim 4; SEQ ID No 14877; 51000; English.  
 XX  
 SS The present invention relates to single exon nucleic acid probes for  
 CC measuring human gene expression in a sample derived from human heart. The  
 CC present sequence is one such probe. The probes may be used for  
 CC predicting, measuring and displaying gene expression in samples derived  
 CC from the human heart via microarrays. By measuring gene expression, the  
 CC probes are useful for predicting, diagnosing, grading, staging,  
 CC monitoring and proposing diseases of the human heart and vascular system  
 CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and  
 CC congenital heart disease.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 77 BP; 8 A; 15 C; 17 G; 37 T; 0 other;

Query Match 82.2%; Score 14.8; DB 22; Length 77;  
 Best Local Similarity 88.9%; Pred. No. 4e-02;  
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCGGCTCCACTAGG 18  
 D6 1 ATGGCGGCTCCACTAGG 18  
 ||||| ||||| |||||

RESULT 13  
 AAC16683  
 ID AAC16683 standard; cDNA; 321 BP.

XX AAC16681;  
 XX AC  
 XX DT 06-OCT-2000 (first entry)  
 XX DE  
 XX KW Human secreted protein 5' EST, SEQ ID NO: 20758.  
 XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 XX KW gene therapy; chromosome mapping; ss.  
 XX OS Homo sapiens.  
 XX PV EP1033401-A2.  
 XX PN  
 XX PD 06-SEP-2000.  
 XX PF 21-FEB-2000; 2000EP-020610.  
 XX PR 26-FEB 1995; 99US-0122487.  
 XX PA (GSE1) GENSET.  
 XX PI Dumas Wilma Ekman C, Ductuit A, Giridharu A;  
 XX DR WPI: 2000 500181/45.  
 XX PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures  
 XX  
 XX Claim 1: SEQ ID 20758; 71bp, C: 30M; English.  
 XX  
 CC The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. No GFP has yet been conclusively  
 CC identified within the present sequence. The 5' ESTs were prepared from  
 CC total human RNAs or polyA+ RNAs derived from 10 different tissues. EST  
 CC sequences usually correspond mainly to the 5' untranslated region (UTR)  
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
 CC derived from the 5' ends of mRNAs and even in those cases where longer  
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 CC They are used to obtain upstream regulatory sequences and to design  
 CC expression and secretion vectors.  
 XX  
 XX Sequence 1: 3 BP; 107 A; 66 C; 73 G; 77 T; 0 other;  
 XX  
 Query Match 92.2%; Score 14.6; DB 21; Length 123;  
 Best Local Similarity 88.9%; Pred. No. 4.3e+22;  
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 ATGGCGCTCCATTAGG 18  
 |||||  
 DB 211 ATGGCAGCTCCATTAGG 248  
 RESULT 14  
 ABA46358  
 ID ABA46358 standard; DNA: 446 BP.  
 XX  
 XX ABA46358;  
 XX AC  
 XX DT 01-FEB-2000 (first entry)  
 XX DE Human breast cell single exon nucleic acid probe #1053.  
 XX KW Human; microarray; single exon probe; gene expression; breast;  
 XX KW disease; cancer; ss.  
 XX OS Homo sapiens.  
 XX PN W0200157274 A2.

PN W0200157274 A2.  
 XX  
 XX PD 09-AUG-2001.  
 XX PF 30-JAN-2001; 2001KW 1007167.  
 XX PR 04-FEB-2000; 2000US-0180412.  
 XX PR 26-MAY-2000; 2000US-0007156.  
 XX PR 30-JUN-2000; 2000US-0604408.  
 XX PR 03-AUG-2000; 2000US-0637166.  
 XX PR 21-SEP-2000; 2000US-0134687.  
 XX PR 27-SEP-2000; 2000US-0530159.  
 XX PR 04-OCT-2000; 2000JP 0144564.  
 XX PA MOLECULAR DYNAMICS INC.  
 XX  
 XX Penn SU, Hanzel UK, Van W, Rank DP;  
 XX WPI: 2001 496931/44.  
 XX  
 XX New spatially addressable set of single exon nucleic acid probes,  
 XX useful for measuring gene expression in a sample derived from human  
 XX breast, comprising a set of single exon nucleic acid probes  
 XX Claim 1: SEQ ID 496931; 42bp, C: 20M; English.  
 XX  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human breast and BT 475 cells. The method involves contacting  
 CC the probes with a collection of detectably labeled nucleic acids  
 CC derived from mRNA of an in breast, and then measuring the label  
 CC bound to each probe of the microarray. The probes are useful for  
 CC verifying the expression of regions of genomic DNA predicted to  
 CC encode proteins. They are useful for gene discovery, and for  
 CC determining predisposition and/or forecasting breast disease  
 CC expression analysis is useful for assessing the toxicity of chemical  
 CC agents on cells. The microarray of this invention presents a far greater  
 CC diversity of probes for measuring gene expression, with far less time  
 CC than expressed sequence tag microarrays. The method is suitable for  
 CC rapid production of functional information from genomic sequences. The  
 CC present sequence is a single exon nucleic acid probe of the invention.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, it was obtained in electronic format directly  
 CC from WIPAC at [http://wipac101/pmc/pmc/publib/est\\_sequences](http://wipac101/pmc/pmc/publib/est_sequences).  
 XX  
 XX Sequence 446 BP; 53 A; 145 C; 113 G; 101 T; 0 other;  
 XX  
 Query Match 92.2%; Score 14.8; DB 22; Length 446;  
 Best Local Similarity 88.9%; Pred. No. 4.4e+02;  
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 ATGGCGCTCCATTAGG 18  
 |||||  
 DB 442 ATGGCGCTCCATTAGG 353  
 RESULT 14  
 ABA46356  
 ID ABA46356 standard; DNA: 446 BP.  
 XX  
 XX ABA46356;  
 XX AC  
 XX DT 23-JAN-2000 (first entry)  
 XX DE Probe #502 for gene expression analysis in human breast cell sample.  
 XX KW Human; gene expression; heart; microarray; vascular system; probe;  
 XX KW cardiovascular disease; hypertension; cardiac arrhythmia;  
 XX KW congenital heart disease; ss.  
 XX OS Homo sapiens.  
 XX PN W0200157274 A2.

```

XX 09-AUG-2001.
XX
XX PF 30 JAN 2001; 2001WO-US000666.
XX
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0612366.
XX PR 21-SEP-2000; 2000US-0234587.
XX PR 27-SEP-2000; 2000US-0235359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX
XX (NOTE: MOLECULAR DYNAMICS INC.
XX
XX PI Penn SQ, Hanzel DX, Chen W, Rank DR;
XX
XX DR WF; 2001-488899/53.
XX
XX PI Single exon nucleic acid probes for analyzing gene expression in human
XX hearts -
XX
XX Claim 1: SEQ ID NO 5002; 530pp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart. The
XX present sequence is one such probe. The probes may be used for
XX predicting, measuring and displaying gene expression in samples derived
XX from the human heart via microarrays. By measuring gene expression, the
XX probes are useful for predicting, diagnosing, grading, staging,
XX monitoring and prognosing diseases of the human heart and vascular system
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
XX congenital heart disease.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pat_sequences.
XX
XX Sequence 446 BP; 93 A; 135 C; 117 G; 101 T; 0 other;
XX
XX Query Match 82.2%; Score 14.8; DB 22; Length 446;
XX Best Local Similarity 88.9%; Pred. No. 4.4e-02;
XX Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX : ATGGCGGCGTCCACTAGG 18
XX ||||| ||||
XX 142 ATGGCGGCGTCCACTAGG 159

```

Search completed: October 21, 2003, 00:01:24  
 Job time: 112.3994 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 GenCore Inc.

CM nucleic - nucleic search, using sw mode:

Run on: October 20, 2003, 18:32:49 : Search time 100.466 seconds  
without alignment  
4394.125 Million SWI updates/sec

File: us-10-029-137-8  
Perfect score: 18  
Sequence: 1 atggcggtatcactaggat 18

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 1215223856 residues

Total number of hits satisfying chosen parameters: 48562584

Minimum EB seq length: 0  
Maximum EB seq length: 2000000000

Post processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: EST:

- 1: em\_estb1
- 2: em\_estb1
- 3: em\_estb1
- 4: em\_estb1
- 5: em\_estb1
- 6: em\_estb1
- 7: em\_estb1
- 8: em\_estb1
- 9: gb\_est1
- 10: gb\_est1
- 11: gb\_hrc1
- 12: gb\_hrc1
- 13: gb\_hrc1
- 14: gb\_hrc1
- 15: em\_estb1
- 16: em\_estb1
- 17: em\_gss\_hm1
- 18: em\_gss\_hm1
- 19: em\_gss\_hm1
- 20: em\_gss\_vit1
- 21: em\_gss\_vit1
- 22: em\_gss\_vit1
- 23: em\_gss\_vit1
- 24: em\_gss\_vit1
- 25: em\_gss\_vit1
- 26: em\_gss\_vit1
- 27: em\_gss\_vit1
- 28: gb\_gss1
- 29: gb\_gss1

Pled. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	EB ID	Description
1	18	100.0	344	7 AV686694	AV686694
2	18	100.0	344	9 AV687554	AV687554
3	18	100.0	344	9 AV687795	AV687795
4	18	100.0	344	9 AV688881	AV688881

5	18	100.0	392	9 AV691918	AV691918
6	18	100.0	400	9 AV688878	AV688878
7	18	100.0	630	12 BW789663	BW789663
8	18	100.0	693	10 BW789678	BW789678
9	18	100.0	728	10 BE797754	BE797754
10	18	100.0	744	10 BG721190	BG721190
11	18	100.0	768	10 BF440388	BF440388
12	18	100.0	807	14 CP92984	CP92984
13	18	100.0	913	14 CD55253	CD55253
14	18	100.0	931	10 BG141322	BG141322
15	18	100.0	979	13 BU158756	BU158756
16	18	100.0	988	14 CD189145	CD189145
17	18	100.0	1019	13 BX437420	BX437420
18	18	100.0	1020	13 BX437419	BX437419
19	18	100.0	1077	11 AF118059	AF118059
20	18	100.0	1155	13 BX463257	BX463257
21	18	100.0	1173	13 BX376388	BX376388
22	18	100.0	1177	13 BX376387	BX376387
23	18	100.0	1201	9 AC523437	AC523437
24	18	100.0	1201	13 BX185386	BX185386
25	17	94.4	881	13 BU176649	BU176649
26	16.4	91.1	291	28 BU712039	BU712039
27	16.4	91.1	400	9 AV690802	AV690802
28	16.4	91.1	432	28 AC281251	AC281251
29	16.4	91.1	463	28 AQ469902	AQ469902
30	16.4	91.1	499	9 AU020839	AU020839
31	16.4	91.1	577	29 AG234746	AG234746
32	16.4	91.1	702	28 BZ036744	BZ036744
33	16.4	91.1	860	29 ATH517023	ATH517023
34	16	88.0	243	9 AU234516	AU234516
35	16	88.0	685	13 BU215299	BU215299
36	16	88.0	789	14 CD246323	CD246323
37	16	88.0	795	10 BF183780	BF183780
38	16	88.0	879	13 BQ961555	BQ961555
39	16	88.0	911	13 BQ425134	BQ425134
40	16	88.0	950	13 BU153378	BU153378
41	15.4	85.6	304	13 BQ979503	BQ979503
42	15.4	85.6	345	9 AV688914	AV688914
43	15.4	85.6	523	29 AG249022	AG249022
44	15.4	85.6	525	9 AW825544	AW825544
45	15.4	85.6	539	9 AW786452	AW786452

ALIGNMENTS

RESULT :  
AV686694 544 bp mRNA linear EST 16-JAN-2002  
LOCUS AV686694 GKT Homo sapiens cDNA clone SKCCOF12.5', mRNA sequence.  
DEFINITION AV686694 GKT Homo sapiens cDNA clone SKCCOF12.5', mRNA sequence.  
ACCESSION AV686694  
VERSION AV686694.1 GI:10288557  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 344)  
AUTHORS Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X., Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W., Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X., He,G., Gu,J., Chen,Z. and Han,Z.  
TITLE Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)  
MEDLINE 21625106  
PUBMED 11752456  
COMMENT Contact: Zengqiang Han  
Chinese National Human Genome Center at Shanghai  
311 500 Shouling Road, Zhongjiang Hi Tech Park, Pudong, Shanghai  
201203, P. R. China  
Tel: 86 21-50801919 (ex. 45)

Fax: 86-21-50801922

Email: hanzgachgc.sc.cn

This clone is available at CHGC in Shanghai.

#### FEATURES

Location/Qualifiers

1..344

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="GKCG02"

/tissue type="hepatocellular carcinoma"

/dev stage="Adult"

/lab\_host="SCLR"

/clone\_lib="GKC"

/note="vector: pBluescript sk(-); Site 1: EcoRI; Site 2: XhoI"

#### BASE COUNT

62 a 97 c 85 g 100 t

#### ORIGIN

Query Match 100.0%; Score 18; DB 9; Length 344;

Best Local Similarity 100.0%; Pred. No. 2.4e+22;

Matches 18; Conservations 0; Mismatches 0; Indels 0; Gaps 0;

#### QY

1 ATGCGGCTTCATCATG 18

#### DB

8 ATGCGGCTTCATCATG 26

#### RESULT 3

AV687554

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

CONTACT

CHINESE NATIONAL HUMAN GENOME CENTER AT SHANGHAI

351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai

201203, P. R. China

Tel: 86-21-50801919 (ex. 45)

Fax: 86-21-50801922

Email: hanzgachgc.sc.cn

This clone is available at CHGC in Shanghai.

Location/Qualifiers

1..344

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="GKCG02"

/tissue type="hepatocellular carcinoma"

/dev stage="Adult"

/lab\_host="SCLR"

/clone\_lib="GKC"

/note="vector: pBluescript sk(-); Site 1: EcoRI; Site 2: XhoI"

#### BASE COUNT

62 a 97 c 85 g 100 t

#### ORIGIN

Query Match 100.0%; Score 18; DB 9; Length 344;

Best Local Similarity 100.0%; Pred. No. 2.4e+22;

Matches 18; Conservations 0; Mismatches 0; Indels 0; Gaps 0;

#### QY

1 ATGCGGCTTCATCATG 18

#### DB

8 ATGCGGCTTCATCATG 26

Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 18; Conservations 0; Mismatches 0; Indels 0; Gaps 0;

#### QY

1 ATGCGGCTTCATCATG 18

#### DB

8 ATGCGGCTTCATCATG 26

#### RESULT 3

AV687554

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

CONTACT

CHINESE NATIONAL HUMAN GENOME CENTER AT SHANGHAI

351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai

201203, P. R. China

Tel: 86-21-50801919 (ex. 45)

Fax: 86-21-50801922

Email: hanzgachgc.sc.cn

This clone is available at CHGC in Shanghai.

Location/Qualifiers

1..344

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="GKCG02"

/tissue type="hepatocellular carcinoma"

/dev stage="Adult"

/lab\_host="SCLR"

/clone\_lib="GKC"

/note="vector: pBluescript sk(-); Site 1: EcoRI; Site 2: XhoI"

#### BASE COUNT

62 a 97 c 85 g 100 t

#### ORIGIN

Query Match 100.0%; Score 18; DB 9; Length 344;

Best Local Similarity 100.0%; Pred. No. 2.4e+22;

Matches 18; Conservations 0; Mismatches 0; Indels 0; Gaps 0;

#### QY

1 ATGCGGCTTCATCATG 18

#### DB

8 ATGCGGCTTCATCATG 26

#### RESULT 4

AV688881

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

CONTACT

CHINESE NATIONAL HUMAN GENOME CENTER AT SHANGHAI

351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai

201203, P. R. China

Tel: 86-21-50801919 (ex. 45)

Fax: 86-21-50801922

Email: hanzgachgc.sc.cn

This clone is available at CHGC in Shanghai.

Location/Qualifiers

1..344

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="GKCG02"

/tissue type="hepatocellular carcinoma"

/dev stage="Adult"

/lab\_host="SCLR"

/clone\_lib="GKC"

/note="vector: pBluescript sk(-); Site 1: EcoRI; Site 2: XhoI"

REFERENCE  
AUTHORS

1 (bases 1 to 344)  
 Xu,X., Huang,J., Xu,Z., Qian,H., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,  
 Xiao,H., Qiu,G., Lin,F., Huang,G., Cheng,Z., Li,N., Du,J., Hu,W.,  
 Shen,K., Lu,G., Fu,G., Zhong,X., Xu,S., Gu,W., Huang,W., Zhao,X.,  
 Hu,G., Gu,J., Chen,Z. and Han,Z.

## TITLE

Insight into hepatocellular carcinogenesis at transcriptome level  
 by comparing gene expression profiles of hepatocellular carcinoma  
 with those of corresponding noncancerous liver

JOURNAL  
MEDLINE  
PubMed

Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)  
 21625106  
 11752456

## COMMENT

Contact: Zeguang Han  
 Chinese National Human Genome Center at Shanghai  
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
 201203, P. R. China  
 Tel: 86-21-50801929 (ex.45)  
 Fax: 86-21-50801922  
 Email: hanzg@cnhg.sh.cn

## FEATURES

This clone is available at CHGC in Shanghai.

## Location/Qualifiers

1..344  
 /organism="Homo sapiens"  
 /mol\_type="cDNA"  
 /db\_xref="taxon:9606"  
 /clone="GKCGH05"  
 /tissue\_type="hepatocellular carcinoma"  
 /dev\_stage="Adult"  
 /lab\_host="SOLR"  
 /clone\_lib="GKC"  
 /note="Vector: pBluescript sk(-); Site\_1: EcoRI; Site\_2:  
 XhoI"

## BASE COUNT

62 a 97 c 85 g 100 t

## ORIGIN

Query Match 100.0%; Score 18; DB 9; Length 344;

Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## CY

1 ATGGCGGCTCCACTAGG 18

||||| ||||| |||||

8 ATGGCGGCTCCACTAGG 25

## RESULT 4

## AV691918

LOCUS AV691918 GRC Homo sapiens cDNA clone GRC91918, mRNA sequence,  
 EST.  
 DEFINITION AV691918  
 ACCESSION AV691918  
 VERSION AV691918.1 GI:10294741  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE

## 1 (bases 1 to 392)

## AUTHORS

Xu,X., Huang,J., Xu,Z., Qian,H., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,  
 Xiao,H., Qiu,G., Lin,F., Huang,G., Cheng,Z., Li,N., Du,J., Hu,W.,  
 Shen,K., Lu,G., Fu,G., Zhong,X., Xu,S., Gu,W., Huang,W., Zhao,X.,  
 Hu,G., Gu,J., Chen,Z. and Han,Z.

## TITLE

Insight into hepatocellular carcinogenesis at transcriptome level  
 by comparing gene expression profiles of hepatocellular carcinoma  
 with those of corresponding noncancerous liver

## JOURNAL

## MEDLINE

## PubMed

Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)

## COMMENT

Contact: Zeguang Han  
 Chinese National Human Genome Center at Shanghai  
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
 201203, P. R. China  
 Tel: 86-21-50801929 (ex.45)  
 Fax: 86-21-50801922  
 Email: hanzg@cnhg.sh.cn

This clone is available at CHGC in Shanghai.

## FEATURES

## source

Location/Qualifiers  
 1..392  
 /organism="Homo sapiens"  
 /mol\_type="cDNA"  
 /db\_xref="taxon:9606"  
 /clone="GKCGH05"  
 /tissue\_type="hepatocellular carcinoma"  
 /dev\_stage="Adult"  
 /lab\_host="SOLR"  
 /clone\_lib="GKC"  
 /note="Vector: pBluescript sk(-); Site\_1: EcoRI; Site\_2:  
 XhoI"

## BASE COUNT

73 a 104 c 112 g 103 t

## ORIGIN

Query Match 100.0%; Score 18; DB 9; Length 392;

Best Local Similarity 100.0%; Pred. No. 2.5e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## CY

1 ATGGCGGCTCCACTAGG 18

||||| ||||| |||||

8 ATGGCGGCTCCACTAGG 25

## RESULT 6

## AV698878

LOCUS AV698878 GRC Homo sapiens cDNA clone GRC98878, mRNA sequence,  
 EST.  
 DEFINITION AV698878  
 ACCESSION AV698878  
 VERSION AV698878.1 GI:10294741  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 400)  
 Xu,X., Huang,J., Xu,Z., Qian,H., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,  
 Xiao,H., Qiu,G., Lin,F., Huang,G., Cheng,Z., Li,N., Du,J., Hu,W.,  
 Shen,K., Lu,G., Fu,G., Zhong,X., Xu,S., Gu,W., Huang,W., Zhao,X.,  
 Hu,G., Gu,J., Chen,Z. and Han,Z.

Insight into hepatocellular carcinogenesis at transcriptome level  
 by comparing gene expression profiles of hepatocellular carcinoma  
 with those of corresponding noncancerous liver

Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)

Contact: Zeguang Han  
 Chinese National Human Genome Center at Shanghai  
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai  
 201203, P. R. China  
 Tel: 86-21-50801929 (ex.45)  
 Fax: 86-21-50801922  
 Email: hanzg@cnhg.sh.cn

This clone is available at CHGC in Shanghai.

Location/Qualifiers  
 1..400  
 /organism="Homo sapiens"  
 /mol\_type="cDNA"  
 /db\_xref="taxon:9606"  
 /clone="GKCGH05"  
 /tissue\_type="hepatocellular carcinoma"  
 /dev\_stage="Adult"  
 /lab\_host="SOLR"  
 /clone\_lib="GKC"  
 /note="Vector: pBluescript sk(-); Site\_1: EcoRI; Site\_2:  
 XhoI"

BASE COUNT 64 a 99 c 123 g 113 t

Query Match 100.0%; Score 18; DB 9; Length 400;

Best Local Similarity 100.0%; Pred. No. 2.5e+02;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;





cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (HML)  
 cDNA Sequencing by: Incyte Genomics, Inc.  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/ILN at: [image.lln.gov](http://image.lln.gov)  
 Plate: LUM781 row: C column: 18  
 High quality sequence start: 26  
 High quality sequence stop: 721.  
 Location/Qualifiers

## FEATURES

source

1..728

/organism="Homo sapiens"

/mol\_type="cDNA"

/db\_xref="taxon:9606"

/clone="IMAGE:494613"

/issue\_type="small cell carcinoma"

/cell\_line="MGC3"

/lab\_host="DH12B (phage resistant)"

/clone\_lib="NIH\_MGC\_7"

/notes="Organ: lung; Vector: pOTB7; Site 1: XhoI; Site 2:

EcoRI; cDNA made by oligo-dT priming. Directionally

cloned into pCDR1/XhoI sites using the following 5'

adapter: GGCACGAGTC. Size-selected >50bp for average

insert size 1.8kb. Library constructed by Ling Hong in

the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)."

183 a 177 c 193 g 175 t

BASE COUNT

ORIGIN

Query Match 100.0% Score 18; DB 10; Length 728;

Best Local Similarity 100.0% Pred. No. 2.9e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGCGCTCCACTAGG 18

Db 42 ATGGCGCTCCACTAGG 59

## RESULT 10

RG723190

LOCUS

DEFINITION 602690755F1 NIH\_MGC\_97 Homo sapiens cDNA clone IMAGE:482313 5',

RNA sequence.

RG723190

VERSION RG723190.1 GI:1460377

KEYWORDS EST.

SOURCE Homo sapiens

ORGANISM Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 744)

NIH-MGC <http://imgc.ncl.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs@mail.nih.gov](mailto:cgapbs@mail.nih.gov)

Tissue Procurement: Miklos Palkovics, M.D., Ph.D.

cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki

Toshiyuki and Piero Carninci (RIKEN)

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LJNL)

cDNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/ILN at:

<http://image.lln.gov>

Plate: JAM10732 row: 1 column: 22

High quality sequence stop: 741.

Location/Qualifiers

1..744

/organism="Homo sapiens"

/mol\_type="cDNA"

/db\_xref="taxon:9606"

/clone="IMAGE:482313"

/lab\_host="DH10B"

## FEATURES

source

1..744

/organism="Homo sapiens"

/mol\_type="cDNA"

/db\_xref="taxon:9606"

/clone="IMAGE:482313"

/lab\_host="DH10B"

## Query Match

100.0% Score 18; DB 10; Length 788;

Best Local Similarity 100.0% Pred. No. 2.9e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGCGCTCCACTAGG 18

Db 12 ATGGCGCTCCACTAGG 29

## RESULT 12

/clone\_lib="NIH\_MGC\_97"

/notes="Organ: testis; Vector: pBluescriptR (modified

pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag

); Oligo-87 primed using primer 5'-TTTTTTTCTTTTNN-3',

size-selected for average insert size 2.2 kb and

normalized to 400 5. This is a primary library enriched

for full-length clones and constructed using the

Cap trapper method (Carninci, in preparation). Library

constructed by M. Brownstein (NIH/NHGRI, National

Institutes of Health). Note: this is a NIH\_MGC library."

188 a 179 c 188 g 187 t

BASE COUNT

ORIGIN

Query Match 100.0% Score 18; DB 10; Length 744;

Best Local Similarity 100.0% Pred. No. 2.9e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGCGCTCCACTAGG 18

Db 15 ATGGCGCTCCACTAGG 32

## RESULT 11

BF440088

LOCUS

DEFINITION BF440088 Homo sapiens cDNA, mRNA sequence.

BF440088

VERSION BF440088.1 GI:11467985

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 788)

Nguyen M.H. and Liu C.H.

Identification of Genes in HUVECs

Unpublished

Contact: Mai H. Nguyen, Canhui Liu

Dr. Mai Nguyen

University of California Los Angeles

54117 CHS, Surgery Oncology, UCLA, 10833 LeConte Ave., Los Angeles,

CA 90095, USA

Tel: 310 267 1963

Fax: 310 825 7575

Email: [mai.nguyen@mednet.ucla.edu](mailto:mai.nguyen@mednet.ucla.edu); [liuch88@yahoo.com](mailto:liuch88@yahoo.com)

PCR Primers

FORWARD: T7

BACKWARD: T3

Insert Length: 788 Std Error: 0.00

Seq primer: T7 and T3

POLYA=yes.

Location/Qualifiers

1..788

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/cell\_line="Human Umbilical Vein Endothelial Cell"

/clone\_lib="HUVEC cDNA Library"

/notes="Vector: Lambda UniZAP Vector; Site 1: EcoRI;

Site 2: XhoI"

216 a 179 c 194 g 199 t

BASE COUNT

ORIGIN

## Query Match

100.0% Score 18; DB 10; Length 788;

Best Local Similarity 100.0% Pred. No. 2.9e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGCGCTCCACTAGG 18

Db 12 ATGGCGCTCCACTAGG 29



Search completed: October 21, 2023, 02:39:45  
Cdb time : 133.466 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

CM nuclei: nucleic search, using sw model  
Run on: October 20, 2003, 17:12:14 Search time 97.6334 seconds  
(without alignment)  
7537.592 Million cell updates/sec

Title: US-10-029-137-8  
Perfect score: 18  
Sequence: 1 atggcggtctactaggg 18

Scoring table: IDENTITY NUC  
Gap: 10.0, Gapext 1.0

Searched: 2888711 seqs, 2045441386 residues  
Total number of hits satisfying chosen parameters: 5077422

Minimum DB seq length: 9  
Maximum DB seq length: 2000000000  
Post processing: Minimum Match 8%  
Maximum Match 100%  
Listing first 45 summaries

Database: GenBank

- 1: gb\_ba.\*
- 2: gb\_hg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pa.\*
- 7: gb\_ph.\*
- 8: gb\_pi.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_st.\*
- 12: gb\_sv.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_fun.\*
- 18: em\_in.\*
- 19: em\_ru.\*
- 20: em\_ro.\*
- 21: em\_ov.\*
- 22: em\_pa.\*
- 23: em\_pi.\*
- 24: em\_ph.\*
- 25: em\_pi.\*
- 26: em\_ro.\*
- 27: em\_st.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_hg\_hum.\*
- 31: em\_hg\_inv.\*
- 32: em\_hg\_other.\*
- 33: em\_hg\_mus.\*
- 34: em\_hg\_pln.\*
- 35: em\_hg\_rod.\*
- 36: em\_hg\_mam.\*
- 37: em\_hg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_hgoc\_hum.\*
- 40: em\_hgoc\_mus.\*
- 41: em\_hgoc\_other.\*

Prod. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	100.0	465	B0025539	Sequence
2	18	100.0	467	B0108605	BD:08605 EST and e
3	18	100.0	783	A3112640	A3112680 Homo sapi
4	18	100.0	822	B0015161	B0015361 Human pro
5	18	100.0	982	A0014917	A0014817 Sequence
6	18	100.0	1071	A0017517	A0017517 Sequence
7	18	100.0	1071	B0135207	B0135207 Human nuc
8	18	100.0	1083	B0011936	B0011936 Homo sapi
9	18	100.0	1288	A0015050	A0015050 Sequence
10	18	100.0	1363	A0015829	A0015829 Homo sapi
11	18	100.0	1363	A0015829	A0015829 Homo sapi
12	18	100.0	179310	A0006160	A0006160 Homo sapi
13	18	100.0	201533	A0006160	A0006160 Homo sapi
14	17	94.4	212115	A0016731	AL173248 Mouse DNA
15	16.4	91.1	33464	B0255924	B0255924 Human DNA
16	16.4	91.1	57561	A0013682	A0013682 Homo sapi
17	16.4	91.1	64281	A0013720	A0013720 Human DNA
18	16.4	91.1	79079	A0013327	A0013027 Homo sapi
19	16.4	91.1	106320	A0002304	A0002304 Genomic s
20	16.4	91.1	109171	A0002328	A0002328 Genomic s
21	16.4	91.1	110000	A0035694	Continuation (3 of
22	16.4	91.1	111378	A00136123	A00136123 Human DNA
23	16.4	91.1	131145	A0006923	A0006923 Homo sapi
24	16.4	91.1	137473	A0015944	A0015944 Homo sapi
25	16.4	91.1	145080	A0014605	A0014605 Mus muscu
26	16.4	91.1	151528	A0024031	A0024031 Homo sapi
27	16.4	91.1	168513	A0011982	A0011982 Homo sapi
28	16.4	91.1	173117	A00157885	A00157885 Homo sapi
29	16.4	91.1	194476	A0015722	A0015722 Mus muscu
30	16.4	91.1	215485	A0013681	A0013681 Homo sapi
31	16.4	91.1	278593	A0011376	A0011376 Rattus no
32	16	89.9	244	B0016709	B0016709 Silkworm
33	16	89.9	154954	A0022507	A0022507 Homo sapi
34	16	89.9	164812	A0078920	A0078920 Homo sapi
35	16	89.9	186361	A0079159	A0079159 Homo sapi
36	16	89.9	193008	A0026676	A0026676 Homo sapi
37	16	89.9	201022	A0020148	A0020148 Mus muscu
38	15.4	85.6	2139	A0046521	A0046521 Macaca fa
39	15.4	85.6	2454	A0054141	A0054141 Homo sapi
40	15.4	85.6	3663	A0009414	A0009414 P. paucicell
41	15.4	85.6	3704	A00393614	A00393614 Staurasir
42	15.4	85.6	41367	HSU143C10	Z66326 Human DNA s
43	15.4	85.6	79170	A0015396	A0015396 Drosophila
44	15.4	85.6	104737	CNS08095	AL1752382 Oryza sat
45	15.4	85.6	119004	A0010230	A0010230 Homo sapi

ALIGNMENTS

RESULT 1  
B0025539  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Sequence tag and encoded human protein.  
465 bp DNA linear PAT 27-AUG 2002  
B0025539, 1, GI:22566762  
JP 2001269182-A/1785.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 465)  
Edwards, J. B. D. M., Duclair, E. and Jordan, J. Y.  
Sequence tag and encoded human protein  
Patent: JP 2001269182-A 1785 02-OCT-2001

```

GENSET
CS Homo sapiens (human)
PN JP 2002010789-A/1785
PD 02-OCT-2000
PF 24-FEB-2000 JP 2000118773
PR 26-FEB-1999 US 60/122487
P1 JEAN BAPTISTE DUMAS MILNE EDWARDS, EIMERIC DUCLAIR, JEAN YVES
P2 JORDANO
PC C12N15/09, C07K14/435, C07K16/18, C12N1/15, C12N1/19, C12N1/21, PC
C12N5/10
PC C12P21/02, C12P21/08, C12Q1/68, C06F17/30, C12N15/00, C12N5/09, PC
C06F15/43
CC
FH Key Location/Qualifiers
FT CDS 20..467
FEATURES
source
Location/Qualifiers
1..465
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 116 a 113 c 122 g 108 t 6 others
ORIGIN
Query Match 100.0%; Score 18; DB 6; Length 467;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCGGCTCCACTAGG 18
DB 20 ATGGCGGCTCCACTAGG 37

RESULT 2
BD108605
LOCUS EST and encoded human protein.
DEFINITION BD108605
ACCESSION BD108605.1 37:23203423
VERSION JP 2002010789-A/1785
KEYWORDS Homo sapiens (human);
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 467)
AUTHORS Edwards J B D M., Robert S. and Giordano, J F.
TITLE EST and encoded human protein
JOURNAL Patent: JP 2002010789-A 692 15-JAN 2002;
GENSET CORP
COMMENT
OS Homo sapiens (human);
PN JP 2002010789-A/682
PD 15-JAN-2002
PF 07-AUG-2000 JP 2000280989
PR 25-AUG-1999 US 60/147499
P1 JEAN BAPTISTE DUMAS MILNE EDWARDS, SEVELIN JOBERT, JEAN YVES P1
GIORDANO
PC C12N15/09, C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, PC
C12N1/21
PC C12N5/10, C12P21/02, C12P21/08, C12Q1/68, C12N15/00, C12N5/00, PC
C12N15/00
CC EST and encoded human protein
FH Key Location/Qualifiers
FT CDS 15..467
FEATURES
source
Location/Qualifiers
1..467
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 121 a 113 c 123 g 107 t 3 others
ORIGIN
Query Match 100.0%; Score 18; DB 6; Length 467;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GENSET
CS Homo sapiens (human)
PN JP 2002010789-A/1785
PD 02-OCT-2000
PF 24-FEB-2000 JP 2000118773
PR 26-FEB-1999 US 60/122487
P1 JEAN BAPTISTE DUMAS MILNE EDWARDS, EIMERIC DUCLAIR, JEAN YVES
P2 JORDANO
PC C12N15/09, C07K14/435, C07K16/18, C12N1/15, C12N1/19, C12N1/21, PC
C12N5/10
PC C12P21/02, C12P21/08, C12Q1/68, C06F17/30, C12N15/00, C12N5/09, PC
C06F15/43
CC
FH Key Location/Qualifiers
FT CDS 20..467
FEATURES
source
Location/Qualifiers
1..465
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 116 a 113 c 122 g 108 t 6 others
ORIGIN
Query Match 100.0%; Score 18; DB 6; Length 467;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCGGCTCCACTAGG 18
DB 20 ATGGCGGCTCCACTAGG 37

RESULT 2
BD108605
LOCUS EST and encoded human protein.
DEFINITION BD108605
ACCESSION BD108605.1 37:23203423
VERSION JP 2002010789-A/1785
KEYWORDS Homo sapiens (human);
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 467)
AUTHORS Edwards J B D M., Robert S. and Giordano, J F.
TITLE EST and encoded human protein
JOURNAL Patent: JP 2002010789-A 692 15-JAN 2002;
GENSET CORP
COMMENT
OS Homo sapiens (human);
PN JP 2002010789-A/682
PD 15-JAN-2002
PF 07-AUG-2000 JP 2000280989
PR 25-AUG-1999 US 60/147499
P1 JEAN BAPTISTE DUMAS MILNE EDWARDS, SEVELIN JOBERT, JEAN YVES P1
GIORDANO
PC C12N15/09, C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, PC
C12N1/21
PC C12N5/10, C12P21/02, C12P21/08, C12Q1/68, C12N15/00, C12N5/00, PC
C12N15/00
CC EST and encoded human protein
FH Key Location/Qualifiers
FT CDS 15..467
FEATURES
source
Location/Qualifiers
1..467
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 121 a 113 c 123 g 107 t 3 others
ORIGIN
Query Match 100.0%; Score 18; DB 6; Length 467;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ATGGCGGCTCCACTAGG 18
DB 15 ATGGCGGCTCCACTAGG 32

RESULT 3
AF317680
LOCUS Homo sapiens tumor-related protein mRNA, complete cds.
DEFINITION AF317680
ACCESSION AF317680.1 3111021849
VERSION AF317680.1 3111021849
KEYWORDS Homo sapiens (human);
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 789)
AUTHORS Mai K. and Liang
TITLE A gene from human
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 789)
AUTHORS Mai K. and Liang
TITLE Direct Submission
JOURNAL Submitted (02-NOV-2002) Sargical Oncology, University of
California, Los Angeles, 10833 Le Conte Ave., Los Angeles, CA
90095, USA
FEATURES
source
Location/Qualifiers
1..789
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
12..591
/codon_start=1
/product="tumor-related protein"
/protein_id="AAK11561.1"
/db_xref="GI:11021850"
/translation="MAAPLQEMESDPFGPPQAPPLPGCASLGAAPAPAPSSSTL
VPLESSEKACASLSDVEYKSTGEEIKTQVDCIOKPLDIAROTECFICARLQI
SVQRPQVTKSDNSLELNELQKCALVQKLTLPKRWQQLVEDINVOHKKKIDITVGS
LADTASGQKQETSEANSSKQRQAYEWADA"
BASE COUNT 216 a 170 c 194 g 195 t 2 others
ORIGIN
Query Match 100.0%; Score 18; DB 9; Length 789;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCGGCTCCACTAGG 18
DB 12 ATGGCGGCTCCACTAGG 29

RESULT 4
BD015361
LOCUS Human protein and cDNA.
DEFINITION BD015361
ACCESSION BD015361.1 6122556499
VERSION JP 200218564-A/7
KEYWORDS JP 200218564-A/7
SOURCE Homo sapiens (human);
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 622)
AUTHORS Kato S. and Siek, M
TITLE Human Protein and cDNA.4
JOURNAL Patent: JP 200218564-A 7 14-AUG-2001;
JAPAN: SCIENCE AND TECHNOLOGY CORP
COMMENT
OS Homo sapiens (human);
PN JP 200218564-A/7
PD 14-AUG-2001
PF 08-FEB-2000 JP 2000031062

```

```

PI SEIJI KATO,MIHORO SAEKI
PC C12N15/09,C07K14/403,C07K16/19,C12N1/21,C12N1/19,C12N1/21,C12N1/21, PC
C12N5/10//
PC C12P21/08,C12N15/09,C12N5/00
CC
FH
PI Key Location/Qualifiers
PI CDS Location/Qualifiers
1..822
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
210 a 186 c 199 g 227 t
BASE COUNT
ORIGIN
Query Match 100.0%; Score 18; DB 6; Length 822;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCGGCTCCACTAGG 18
|||||
Db 10 ATGGCGGCTCCACTAGG 27
RESULT 5
AX014817 882 bp DNA linear PAT 07-SEP-2000
LOCUS
DEFINITION
Sequence 3 from Patent WO951040.
ACCESSION
AX014817
VERSION
AX014817.1 GI:110041084
KEYWORDS
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS
Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Pilarsky,C.
TITLE
Human nucleic acid sequences from ovarian tumour tissue
JOURNAL
Patent: WO 9953040-A 3 21-OCT-1999;
SCHMITT ARVIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILARSKY CHRISTIAN (DE)
(UE); PILARSKY CHRISTIAN (DE)
FEATURES
Location/Qualifiers
1..882
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
238 a 190 c 205 g 245 t
BASE COUNT
ORIGIN
Query Match 100.0%; Score 18; DB 6; Length 882;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCGGCTCCACTAGG 18
|||||
Db 10 ATGGCGGCTCCACTAGG 27
RESULT 6
AX017517 1071 bp DNA linear PAT 07-SEP-2000
LOCUS
DEFINITION
Sequence 65 from Patent WO9947655.
ACCESSION
AX017517
VERSION
AX017517.1 GI:110042314
KEYWORDS
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS
Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and

```

```

Pilarsky,C.
Human nucleic acid sequences from normal breast tissue
Patent: WO 9947655-A 65 23-SEP-1999;
SCHMITT ARVIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILARSKY CHRISTIAN (DE)
(UE); PILARSKY CHRISTIAN (DE)
FEATURES
Location/Qualifiers
1..1071
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
281 a 218 c 255 g 317 t
BASE COUNT
ORIGIN
Query Match 100.0%; Score 18; DB 6; Length 1071;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCGGCTCCACTAGG 18
|||||
Db 10 ATGGCGGCTCCACTAGG 27
RESULT 7
BD135207 1071 bp DNA linear PAT 18-SEP-2002
LOCUS
DEFINITION
Human nucleic acid sequence originating in normal mammary tissue.
ACCESSION
BD135207
VERSION
BD135207.1 GI:23230152
KEYWORDS
JP 2002506639-A/54;
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS
Speft,T., Hinzmann,B., Schmitt,A., Pirarski,C., Duhl,E. and
Rosenthal,A.
Human nucleic acid sequence originating in normal mammary tissue
Patent: JP 2002506639-A 54 05-MAR-2002;
METAGEN GESBUJSCHAFT FUER GENOME FORSCHUNG MBH
OS Homo sapiens (human);
PN JP 2002506639-A/54
PD 25-MAR-2002
PF 19-MAR-1999 JP 2000536918
PR 20-MAR-1999 DE 198 13 035.0
PI THOMAS SPEFT,BERND HINZMANN,ARVIN SCHMITT,CHRISTIAN PIRARSKI,
E; EDGAR DUHL,
E; ANDRE ROSENTHAL
PC C12N15/09,A61K48/00,A61P43/00,A61P43/00,C07K14/47,
PC C07K16/18
PC C12N1/21,C12N5/10,C12P21/02,C12Q1/68,G01N33/68/A61K38/00, PC
C12N15/00,
PC C12N5/00,A61K37/02
CC Human nucleic acid sequence originating in normal mammary CC
tissue
FH Key Location/Qualifiers
FH source 1..1071
/organism="Homo sapiens (human)"
FEATURES
Location/Qualifiers
1..1071
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
281 a 218 c 255 g 317 t
BASE COUNT
ORIGIN
Query Match 100.0%; Score 18; DB 6; Length 1071;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCGGCTCCACTAGG 18
|||||
Db 10 ATGGCGGCTCCACTAGG 27

```



```

CDS
13 543
/codon_start=1
/product="tumor angiogenesis marker"
/protein_id="AAK32724.1"
/db_xref="GI:13603460"
/translation="MAAPLGGMFSCQFPBPQANFGI FGQASLLQAAGAPRPSSSTL
VLESSFEACFASLVQDVNQTQCEIRFGVQCICKEFIDIAQTCEFLQKRLCL
SVQKPCVKEQVSELELOKQALVOKHITKPHNCOVLEINVQKXKADIPQGS
LAYEQANIPAPLKP?"
BASE COUNT      410 a   245 c   359 g   199 t
ORIGIN
1
Query Match      100.0%; Score 18; DB 9; Length 1463;
Best local similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCGGCTCCACTAGG 18
|||||
DB 13 ATGGCGGCTCCACTAGG 30

RESULT 13
AF321617
LOCUS Homo sapiens unknown mRNA 3109 bp mRNA 11 May FRI 07-DEC-2003
DEFINITION Homo sapiens unknown mRNA.
ACCESSION AF321617
VERSION AF321617.1 GI:11596417
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 3109)
AUTHORS Nguyen, V.H. and Hsu, C.
TITLE Full-length cDNA for Unknown#3
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3109)
AUTHORS Nguyen, M.H. and Gu, S.
TITLE Direct Submission
JOURNAL Submitted (15-NOV-2003) Sutterland Oncology, University of
California, Los Angeles, 10433 Le Conte Ave., Los Angeles, CA
90095, CA

FEATURES
Source
1..3109
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1755..2109
/codon_start=1
/product="unknown"
/protein_id="AAK3612.1"
/db_xref="GI:11596418"
/translation="MAAPLGGMFSCQFPBPQANFGI FGQASLLQAAGAPRPSSSTL
VLESSFEACFASLVQDVNQTQCEIRFGVQCICKEFIDIAQTCEFLQKRLCL
SVQKPCVKEQVSELELOKQALVOKHITKPHNCOVLEINVQKXKADIPQGS
LAYEQANIPAPLKP?"
BASE COUNT      980 a   503 c   621 g   1005 t
ORIGIN
1
Query Match      100.0%; Score 18; DB 9; Length 2109;
Best local similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCGGCTCCACTAGG 18
|||||
DB 13 ATGGCGGCTCCACTAGG 1776

RESULT 13
AF321617
LOCUS Homo sapiens chromosome 4 clone 1048124 map 416, complete
sequence.
DEFINITION Homo sapiens chromosome 4 clone 1048124 map 416, complete
sequence.
ACCESSION AF321617
VERSION AF321617.1 GI:11596417
KEYWORDS
SOURCE Homo sapiens (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 201533)
AUTHORS Corby, N.
TITLE Direct Submission
JOURNAL Submitted (12-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, Cambs, UK. E-mail enquiries:
humuq@yassanger.ac.uk Clone requests: clonerequests@sanger.ac.uk

CDS
13 543
/codon_start=1
/product="tumor angiogenesis marker"
/protein_id="AAK32724.1"
/db_xref="GI:13603460"
/translation="MAAPLGGMFSCQFPBPQANFGI FGQASLLQAAGAPRPSSSTL
VLESSFEACFASLVQDVNQTQCEIRFGVQCICKEFIDIAQTCEFLQKRLCL
SVQKPCVKEQVSELELOKQALVOKHITKPHNCOVLEINVQKXKADIPQGS
LAYEQANIPAPLKP?"
BASE COUNT      410 a   245 c   359 g   199 t
ORIGIN
1
Query Match      100.0%; Score 18; DB 9; Length 1463;
Best local similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCGGCTCCACTAGG 18
|||||
DB 13 ATGGCGGCTCCACTAGG 30

RESULT 13
AF321617
LOCUS Homo sapiens unknown mRNA 3109 bp mRNA 11 May FRI 07-DEC-2003
DEFINITION Homo sapiens unknown mRNA.
ACCESSION AF321617
VERSION AF321617.1 GI:11596417
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 3109)
AUTHORS Nguyen, V.H. and Hsu, C.
TITLE Full-length cDNA for Unknown#3
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3109)
AUTHORS Nguyen, M.H. and Gu, S.
TITLE Direct Submission
JOURNAL Submitted (15-NOV-2003) Sutterland Oncology, University of
California, Los Angeles, 10433 Le Conte Ave., Los Angeles, CA
90095, CA

FEATURES
Source
1..3109
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1755..2109
/codon_start=1
/product="unknown"
/protein_id="AAK3612.1"
/db_xref="GI:11596418"
/translation="MAAPLGGMFSCQFPBPQANFGI FGQASLLQAAGAPRPSSSTL
VLESSFEACFASLVQDVNQTQCEIRFGVQCICKEFIDIAQTCEFLQKRLCL
SVQKPCVKEQVSELELOKQALVOKHITKPHNCOVLEINVQKXKADIPQGS
LAYEQANIPAPLKP?"
BASE COUNT      980 a   503 c   621 g   1005 t
ORIGIN
1
Query Match      100.0%; Score 18; DB 9; Length 2109;
Best local similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCGGCTCCACTAGG 18
|||||
DB 13 ATGGCGGCTCCACTAGG 1776

RESULT 13
AF321617
LOCUS Homo sapiens chromosome 4 clone 1048124 map 416, complete
sequence.
DEFINITION Homo sapiens chromosome 4 clone 1048124 map 416, complete
sequence.
ACCESSION AF321617
VERSION AF321617.1 GI:11596417
KEYWORDS
SOURCE Homo sapiens (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 201533)
AUTHORS Corby, N.
TITLE Direct Submission
JOURNAL Submitted (12-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, Cambs, UK. E-mail enquiries:
humuq@yassanger.ac.uk Clone requests: clonerequests@sanger.ac.uk

```





```

* be preserved.
* 14404: contig of 14404 bp in length
* 14405: gap of 100 bp
* 14505: contig of 72 bp in length
* 14577: gap of 100 bp
* 14577: contig of 63187 bp in length
* 77863: gap of 100 bp
* 77864: contig of 3576 bp in length
* 81539: gap of 100 bp
* 81540: contig of 3246 bp in length
* 84885: gap of 100 bp
* 84886: gap of 100 bp
* 84985: gap of 100 bp
* 97019: contig of 12354 bp in length
* 98020: gap of 100 bp
* 111968: contig of 13949 bp in length
* 112069: gap of 100 bp
* 132856: contig of 20818 bp in length
* 132887: gap of 100 bp
* 165934: contig of 33008 bp in length
* 165935: gap of 100 bp
* 211359: contig of 45265 bp in length
* 211460: gap of 100 bp
* 211460: contig of 656 bp in length.
Location/Qualifiers
1..212115
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone_lib="RPG1-23 Female Mouse BAC"
1..14404
/notes="assembly_fragment"
clone_end:Sp6
vector_side:left"
misc_feature 14505..14576
/notes="assembly_fragment"
misc_feature 14577..17863
/notes="assembly_fragment"
misc_feature 77964..81539
/notes="assembly_fragment"
misc_feature 81640..84885
/notes="assembly_fragment"
misc_feature 84986..97919
/notes="assembly_fragment"
misc_feature 98020..111968
/notes="assembly_fragment"
misc_feature 112069..132886
/notes="assembly_fragment"
misc_feature 132887..165934
/notes="assembly_fragment"
misc_feature 165935..211359
/notes="assembly_fragment"
misc_feature 211460..212115
/notes="assembly_fragment"
clone_end:T7
vector_side:right"

```

## FEATURES

```

Source
1..212115
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone_lib="RPG1-23 Female Mouse BAC"
1..14404
/notes="assembly_fragment"
clone_end:Sp6
vector_side:left"
misc_feature 14505..14576
/notes="assembly_fragment"
misc_feature 14577..17863
/notes="assembly_fragment"
misc_feature 77964..81539
/notes="assembly_fragment"
misc_feature 81640..84885
/notes="assembly_fragment"
misc_feature 84986..97919
/notes="assembly_fragment"
misc_feature 98020..111968
/notes="assembly_fragment"
misc_feature 112069..132886
/notes="assembly_fragment"
misc_feature 132887..165934
/notes="assembly_fragment"
misc_feature 165935..211359
/notes="assembly_fragment"
misc_feature 211460..212115
/notes="assembly_fragment"
clone_end:T7
vector_side:right"

```

## FEATURES

```

Source
1..14404
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-541119"
/clone_lib="RP11-11.2"

```

```

BASE COUNT 9459 a 6598 c 6584 g 10763 t
ORIGIN

```

```

Query Match 91.1% Score 16.4; DB 9; Length 33404;
Best Local Similarity 94.4%; Pred. No. 2, 1e-02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGGCGCTCCACTAGG 18
|||||
Db 27023 ATGGCGCTCCACTAGG 27036
|||||

```

```

Search completed: October 20, 2003, 23:50:00
Job time : 102.693 secs

```

```

Query Match 94.4% Score 17; DB 2; Length 212115;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCGCTCCACTAGG 17
|||||
Db 202708 ATGGCGCTCCACTAGG 202762
|||||

```

## RESULT 15

```

BX255924/c 33404 bp DNA linear PR: 23-MAR-2003
LOCUS
DEFINITION Human DNA sequence from clone RP11-541119 on chromosome 10,
complete sequence.
ACCESSION BX255924

```

GenScan version 5.1.6  
Copyright (c) 1991 - 2003 CompuLink Ltd.

us nucleic nucleic search, using sw mode:

Run on: October 20, 2003, 21:44:54 : Search time: 41415 seconds  
(without alignment)  
1290.972 Million cell updates/sec

Titles: us-10-029-137-7  
Perfect scores: 18  
Sequence: 1: us-09-029-137-7

Scoring table: 1: us-09-029-137-7

Gapop 10.0, Gapext 1.0

Searches: 569978 seqs, 22069366 residues

Total number of hits satisfying chosen parameters: 113956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Issued Parents NA

- 1: /cgn2\_6/prodata/2/ina/5A COMP seqs
- 2: /cgn2\_6/prodata/2/ina/5P COMP seqs
- 3: /cgn2\_6/prodata/2/ina/5A COMP seqs
- 4: /cgn2\_6/prodata/2/ina/5P COMP seqs
- 5: /cgn2\_6/prodata/2/ina/5A COMP seqs
- 6: /cgn2\_6/prodata/2/ina/5P COMP seqs

Print. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY

Result No.	Score	Match	Length	Parent	Description
1	17.8	94.4	540	US-09-029-137-7	Sequence 1, Appl
2	17.8	94.4	540	US-09-029-137-7	Sequence 55, Appl
3	14.8	82.2	2885	US-09-048-755-2	Sequence 2, Appl
4	14.8	82.2	2885	US-09-050-543-2	Sequence 2, Appl
5	14.8	77.8	2685	US-09-054-856-17	Sequence 5, Appl
6	14.8	77.8	2685	US-09-054-856-17	Sequence 37, Appl
7	14.8	77.8	2685	US-09-054-856-17	Sequence 21, Appl
8	14.8	77.8	2685	US-09-054-856-17	Sequence 5, Appl
9	14.8	77.8	2685	US-09-054-856-17	Sequence 29, Appl
10	14.8	77.8	2685	US-09-054-856-17	Sequence 35, Appl
11	14.8	77.8	2685	US-09-054-856-17	Sequence 19, Appl
12	14.8	77.8	2685	US-09-054-856-17	Sequence 49, Appl
13	14.8	77.8	2685	US-09-054-856-17	Sequence 37, Appl
14	14.8	77.8	2685	US-09-054-856-17	Sequence 17, Appl
15	14.8	77.8	2685	US-09-054-856-17	Sequence 1, Appl
16	14.8	77.8	2685	US-09-054-856-17	Sequence 15, Appl
17	14.8	77.8	2685	US-09-054-856-17	Sequence 15, Appl
18	14.8	77.8	2685	US-09-054-856-17	Sequence 15, Appl
19	14.8	77.8	2685	US-09-054-856-17	Sequence 15, Appl
20	14.8	77.8	2685	US-09-054-856-17	Sequence 15, Appl
21	14.8	77.8	2685	US-09-054-856-17	Sequence 15, Appl
22	14.8	77.8	2685	US-09-054-856-17	Sequence 15, Appl
23	14.8	77.8	2685	US-09-054-856-17	Sequence 15, Appl
24	14.8	77.8	2685	US-09-054-856-17	Sequence 15, Appl
25	14.8	77.8	2685	US-09-054-856-17	Sequence 15, Appl
26	14.8	77.8	2685	US-09-054-856-17	Sequence 15, Appl
27	14.8	77.8	2685	US-09-054-856-17	Sequence 15, Appl

28	13.8	76.7	2312	US-08-102-942A-1	Sequence 1, Appl
29	13.8	76.7	2312	US-09-237-179B-1	Sequence 1, Appl
30	13.8	76.7	3025	US-09-220-132-19	Sequence 1, Appl
31	13.8	76.7	3358	US-09-248-571-2	Sequence 2, Appl
32	13.8	76.7	3358	US-09-553-736-2	Sequence 2, Appl
33	13.8	76.7	3552	US-09-107-532A-3026	Sequence 3026, Appl
34	13.8	76.7	5749	US-08-961-527-94	Sequence 84, Appl
35	13.8	76.7	8460	US-08-469-005A-9	Sequence 9, Appl
36	13.8	76.7	8529	US-09-761-907-1	Sequence 1, Appl
37	13.8	76.7	4257	US-06-311-731A-126	Sequence 126, Appl
38	13.8	76.7	4257	US-09-810-347-3	Sequence 1, Appl
39	13.4	74.4	934	US-09-368-236-1	Sequence 1, Appl
40	13.4	74.4	934	US-09-176-657-6	Sequence 6, Appl
41	13.4	74.4	934	US-09-421-299-6	Sequence 6, Appl
42	13.4	74.4	1350	US-07-792-466-7	Sequence 7, Appl
43	13.4	74.4	1350	US-08-474-633A-1	Sequence 1, Appl
44	13.4	74.4	1350	US-08-737-524B-4	Sequence 4, Appl
45	13.4	74.4	1350	US-08-423-771-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1:  
US-09-109-317-1/5  
Sequence 1, Application US/09109317  
Patent No. 6277970  
GENERAL INFORMATION:  
APPLICANT: Prusiner, Stanley  
APPLICANT: Treibay, Patrick  
APPLICANT: Moore, Richard  
APPLICANT: Westaway, David  
APPLICANT: Hood, Leroy E.  
APPLICANT: Lee, Inyoul  
TITLE OF INVENTION: PrP like Gene  
FILE REFERENCE: 6510-3105  
CURRENT APPLICATION NUMBER: US/09109317  
CURRENT FILING DATE: 1999-05-11  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 1:  
LENGTH: 540  
TYPE: DNA  
ORGANISM: mus musculus  
US-09-109-317-1/5

Query Match: 94.4% Score 17, DS 3, Length 540  
Best local Similarity: 100.0%, Prod. No. 2.8  
Matches: 17, Conservative 0, Mismatches 0, Indels 0, Gaps 0

Cy: TCAGTTCGCTTCAGAG 17  
Db: 301 TCAGTTCGCTTCAGAG 284

RESULT 2:  
US-09-410-112D-653/5  
Sequence 55, Application US/094620312D  
Patent No. 6596f2  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhang, Jie  
APPLICANT: Ren, Feiyang  
APPLICANT: Chen, Rui-Hong  
APPLICANT: Zhao, Qing A.  
APPLICANT: Wehrman, Tom  
APPLICANT: Xue, Aiding J.  
APPLICANT: Yang, Yonghong  
APPLICANT: Wang, Jian Rui  
APPLICANT: Zhang, Ping  
APPLICANT: Ma, Yungqing

1 APPLICANT: Wang, Dantui  
 1 APPLICANT: Wang, Zhiwei  
 1 APPLICANT: John Tillinghast  
 1 APPLICANT: Ertanac, Raduie T.  
 1 TITLE OF INVENTION: NO. 6569662e1 Nucleic Acids and  
 1 Polypeptides  
 1 FILE REFERENCE: 784CIP2B  
 1 CURRENT APPLICATION NUMBER: US/09/620,317C  
 1 CURRENT FILING DATE: 2000-07-19  
 1 PRIOR APPLICATION NUMBER: 09/552,317  
 1 PRIOR FILING DATE: 2000-04-25  
 1 PRIOR APPLICATION NUMBER: 09/488,725  
 1 PRIOR FILING DATE: 2000-01-21  
 1 NUMBER OF SEQ ID NOS: 1155  
 1 SOFTWARE: pt FL\_genes version 1.0  
 1 SEQ ID NO 853  
 1 LENGTH: 2885  
 1 TYPE: DNA  
 1 ORGANISM: Homo sapiens  
 1 FEATURES: CDS  
 1 NAME/KEY: CDS  
 1 LOCATION: 1155  
 1 1444  
 1 US 09 620-312D 853

Query Match: 82.2% Score 14.8; DB 3; Length 2885;  
 Best Local Similarity: 88.9%; Pred. No. 63;  
 Matches: 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCAGTTGGCTTCAGG 18  
 DB 218 TCAGTTGGCTTCAGG 73;

RESULT 3  
 US-08-948 705-2/c  
 1 Sequence 2, Application US/08948705A  
 1 Patent No. 6043054  
 1 GENERAL INFORMATION:  
 1 APPLICANT: Scanlan, Matthew J.  
 1 APPLICANT: Chen, Yao-Tseng  
 1 APPLICANT: Stockert, Elisabeth  
 1 APPLICANT: Old, Lloyd J.  
 1 TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES  
 1 TITLE OF INVENTION: ASSOCIATED WITH COLON CANCER AND METHODS FOR DIAGNOSING AND  
 1 TREATING COLON CANCER  
 1 FILE REFERENCE: LUD-5506-JEL/NDH  
 1 CURRENT APPLICATION NUMBER: US/08/948,705A  
 1 CURRENT FILING DATE: 1997-10-10  
 1 NUMBER OF SEQ ID NOS: 5  
 1 SOFTWARE: FastSeq for Windows Version 3.0  
 1 SEQ ID NO 2  
 1 LENGTH: 2885  
 1 TYPE: DNA  
 1 ORGANISM: Homo Sapiens  
 1 US-08-948 705-2

Query Match: 82.2% Score 14.8; DB 3; Length 2885;  
 Best Local Similarity: 88.9%; Pred. No. 63;  
 Matches: 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCAGTTGGCTTCAGG 18  
 DB 227 TCAGTTGGCTTCAGG 213

RESULT 4  
 US-09-510-543-2/c  
 1 Sequence 2, Application US/09510543  
 1 Patent No. 6517837  
 1 GENERAL INFORMATION:  
 1 APPLICANT: Scanlan, Matthew J.  
 1 APPLICANT: Chen, Yao-Tseng  
 1 APPLICANT: Stockert, Elisabeth

1 APPLICANT: Old, Lloyd J.  
 1 TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ASSOCIATED WITH COLON CANCER AND  
 1 METHODS FOR DIAGNOSING AND TREATING COLON CANCER  
 1 FILE REFERENCE: LUD-6006-JEL/NDH  
 1 CURRENT APPLICATION NUMBER: US/09/510,543  
 1 CURRENT FILING DATE: 2000-02-22  
 1 PRIOR APPLICATION NUMBER: US 07/510,329  
 1 PRIOR FILING DATE: 1998-03-20  
 1 NUMBER OF SEQ ID NOS: 11  
 1 SEQ ID NO 2  
 1 LENGTH: 2885  
 1 TYPE: DNA  
 1 ORGANISM: Homo Sapiens  
 1 US-09-510-543-2

Query Match: 82.2% Score 14.8; DB 4; Length 2885;  
 Best Local Similarity: 88.9%; Pred. No. 63;  
 Matches: 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCAGTTGGCTTCAGG 18  
 DB 227 TCAGTTGGCTTCAGG 213

RESULT 5  
 US-09-854-856-3  
 1 Sequence 53, Application US/09854856  
 1 Patent No. 6541252  
 1 GENERAL INFORMATION:  
 1 APPLICANT: Walker, D. Wayne  
 1 APPLICANT: Halbur, Brian  
 1 APPLICANT: Garibo, Gregory  
 1 APPLICANT: Ertanac, Raduie T.  
 1 TITLE OF INVENTION: NO. 6541252e1 Human Kinases and Polynucleotides  
 1 TITLE OF INVENTION: INCLUDING THE SAME  
 1 FILE REFERENCE: LEX-01-005A  
 1 CURRENT APPLICATION NUMBER: US/09/854,856  
 1 CURRENT FILING DATE: 2001-05-14  
 1 PRIOR APPLICATION NUMBER: US 60/206,015  
 1 PRIOR FILING DATE: 2001-06-14  
 1 NUMBER OF SEQ ID NOS: 14  
 1 SOFTWARE: FastSeq for Windows Version 4.0  
 1 SEQ ID NO 53  
 1 LENGTH: 2885  
 1 TYPE: DNA  
 1 ORGANISM: Homo Sapiens  
 1 US-09-854-856-3

Query Match: 77.04% Score 14; DB 4; Length 2685;  
 Best Local Similarity: 100.0%; Pred. No. 17e+02;  
 Matches: 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAGTTGGCTTCAGG 14  
 DB 2357 TCAGTTGGCTTCAGG 2370

RESULT 6  
 US-09-854-856-37  
 1 Sequence 17, Application US/0984856  
 1 Patent No. 6541252  
 1 GENERAL INFORMATION:  
 1 APPLICANT: Walker, D. Wayne  
 1 APPLICANT: Halbur, Brian  
 1 APPLICANT: Garibo, Gregory  
 1 APPLICANT: Ertanac, Raduie T.  
 1 TITLE OF INVENTION: NO. 6541252e1 Human Kinases and Polynucleotides  
 1 TITLE OF INVENTION: INCLUDING THE SAME

1 FILE REFERENCE: LEX-01-005A  
 1 CURRENT APPLICATION NUMBER: US/09/854,856  
 1 CURRENT FILING DATE: 2001-05-14  
 1 PRIOR APPLICATION NUMBER: US 60/206,015  
 1 PRIOR FILING DATE: 2000-06-19

1 NUMBER OF SEQ ID NOS: 64  
2 SOFTWARE: FastSeq for Windows Version: 4.0  
3 SEQ ID NO: 17  
4 LENGTH: 2169  
5 TYPE: DNA  
6 ORGANISM: Homo sapiens  
US-09-854-854-37

Query Match 77.8%; Score 14; DB 4; Length 2169;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAGTTGGCTTCA 14  
||| |||||  
DB 2441 TCAGTTGGCTTCA 2454

RESULT 7  
US-09-854-854-21  
1 Sequence 21, Application US/09854856  
2 Patent No. 6541252  
3 GENERAL INFORMATION:  
4 APPLICANT: Walke, D. Wade  
5 APPLICANT: Hilburn, Eric  
6 APPLICANT: Donoro, Gregory  
7 APPLICANT: Turner, C. Alexander Jr.  
8 TITLE OF INVENTION: Human Kinase and Polynucleotides  
9 FILE REFERENCE: LEX 0178-JSA  
10 CURRENT APPLICATION NUMBER: US/09/854-856  
11 PRIOR FILING DATE: 2001-05-14  
12 PRCIP APPLICATION NUMBER: US 60/206-015  
13 PRIOR FILING DATE: 2000-05-19  
14 NUMBER OF SEQ ID NOS: 64  
15 SOFTWARE: FastSeq for Windows Version 4.0  
16 SEQ ID NO: 17  
17 LENGTH: 2169  
18 TYPE: DNA  
19 ORGANISM: Homo sapiens  
US-09-854-854-21

Query Match 77.8%; Score 14; DB 4; Length 2865;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAGTTGGCTTCA 14  
||| |||||  
DB 2517 TCAGTTGGCTTCA 2652

RESULT 8  
US-09-854-854-5  
1 Sequence 5, Application US/09854856  
2 Patent No. 6541252  
3 GENERAL INFORMATION:  
4 APPLICANT: Walke, D. Wade  
5 APPLICANT: Hilburn, Eric  
6 APPLICANT: Donoro, Gregory  
7 APPLICANT: Turner, C. Alexander Jr.  
8 TITLE OF INVENTION: Human Kinases and Polynucleotides  
9 FILE REFERENCE: LEX 0178-JSA  
10 CURRENT APPLICATION NUMBER: US/09/854-856  
11 CURRENT FILING DATE: 2001-05-14  
12 PRIOR APPLICATION NUMBER: US 60/206-015  
13 PRIOR FILING DATE: 2000-05-19  
14 NUMBER OF SEQ ID NOS: 64  
15 SOFTWARE: FastSeq for Windows Version 4.0  
16 SEQ ID NO: 5  
17 LENGTH: 2865  
18 TYPE: DNA  
19 ORGANISM: Homo sapiens  
US-09-854-854-5

Query Match 77.8%; Score 14; DB 4; Length 2949;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAGTTGGCTTCA 14  
||| |||||  
DB 2602 TCAGTTGGCTTCA 2634

RESULT 9  
US-09-107-532A-29076  
1 Sequence 230, Application US/09107512A  
2 Patent No. 6581275  
3 GENERAL INFORMATION:  
4 APPLICANT: Lynn A. Doucette-Stamm and David Bush  
5 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
6 ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
7 NUMBER OF SEQUENCES: 7310  
8 CORRESPONDENCE ADDRESS:  
9 ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
10 STREET: 100 Beaver Street  
11 CITY: Waltham  
12 STATE: Massachusetts  
13 COUNTRY: USA  
14 ZIP: 02154  
15 COMPUTER READABLE FORM:  
16 MEDIUM TYPE: CD-ROM ISO9660  
17 OPERATING SYSTEM: <Unknown>  
18 SOFTWARE: ASCII  
19 CURRENT APPLICATION DATA:  
20 APPLICATION NUMBER: US/09/107-532A  
21 FILING DATE: 30-Jun-1998  
22 PRCIP APPLICATION DATA:  
23 APPLICATION NUMBER: 60/085-598  
24 FILING DATE: 14-May-1998  
25 APPLICATION NUMBER: 60/051571  
26 FILING DATE: July 2, 1997  
27 ATTORNEY/AGENT INFORMATION:  
28 NAME: Ariniello, Pamela Deneke  
29 REGISTRATION NUMBER: 40,489  
30 REFERENCE/DOCKET NUMBER: GTC-012  
31 TELECOMMUNICATION INFORMATION:  
32 TELEPHONE: (617) 931-5077  
33 TELEFAX: (617) 931-5077  
34 INFORMATION FOR SEQ ID NO: 290:  
35 SEQUENCE CHARACTERISTICS:  
36 LENGTH: 353; base pairs  
37 TYPE: nucleic acid  
38 STRANDEDNESS: double  
39 TOPOLOGY: circular  
40 MOLECULE TYPE: DNA (genomic)  
41 HYPOTHETICAL: NO  
42 ANTI-SENSE: NO  
43 ORIGINAL SOURCE:  
44 ORGANISM: Enterococcus faecium  
45 FEATURE:  
46 NAME/KEY: misc feature  
47 LOCATION: (8) LOCATION 1...3591  
48 SEQUENCE DESCRIPTION: SEQ ID NO: 290:

QY 1 TCAGTTGGCTTCA 14  
||| |||||  
DB 3397 TCAGTTGGCTTCA 3384

Query Match 77.8%; Score 14; DB 4; Length 3591;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAGTTGGCTTCA 14  
||| |||||  
DB 3397 TCAGTTGGCTTCA 3384

RESULT 10

```

US-09-854-856-51
; Sequence 51, Application US/09854856
; Patent No. 6541252
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Hilbun, Erin
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: LEX-0178-USA
; CURRENT APPLICATION NUMBER: US/09/854,856
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/206,015
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 51
; LENGTH: 6474
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-854-856-51

Query Match 77.8%, Score 14; DB 4; Length 6474;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACGTTGGCTTCA 14
DB 2457 TCACGTTGGCTTCA 2370

RESULT 11
US-09-854-856-35
; Sequence 35, Application US/09854856
; Patent No. 6541252
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Hilbun, Erin
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: LEX-0178-USA
; CURRENT APPLICATION NUMBER: US/09/854,856
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/206,015
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 6558
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-854-856-35

Query Match 77.8%, Score 14; DB 4; Length 6558;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACGTTGGCTTCA 14
DB 244 TCACGTTGGCTTCA 2454

RESULT 12
US-09-854-856-19
; Sequence 19, Application US/09854856
; Patent No. 6541252
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Hilbun, Erin
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: LEX-0178-USA
; CURRENT APPLICATION NUMBER: US/09/854,856
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/206,015
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 6654
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-854-856-19

Query Match 77.8%, Score 14; DB 4; Length 6654;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACGTTGGCTTCA 14
DB 2621 TCACGTTGGCTTCA 2634

RESULT 14
US-09-854-856-44
; Sequence 49, Application US/09854856
; Patent No. 6541252
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Hilbun, Erin
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: LEX-0178-USA
; CURRENT APPLICATION NUMBER: US/09/854,856
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/206,015
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 6738
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-854-856-3

Query Match 77.8%, Score 14; DB 4; Length 6738;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACGTTGGCTTCA 14
DB 2621 TCACGTTGGCTTCA 2634

RESULT 14
US-09-854-856-44
; Sequence 49, Application US/09854856
; Patent No. 6541252
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Hilbun, Erin
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. 6541252el Human Kinases and Polynucleotides
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: LEX-0178-USA
; CURRENT APPLICATION NUMBER: US/09/854,856
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: US 60/206,015
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 6738
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-854-856-3

```

? PRIOR FILING DATE: 2001-05-19  
 ? NUMBER OF SEQ ID NOS: 64  
 ? SOFTWARE: FastSeq for Windows Version 4.0  
 ? SEQ ID NO 49  
 ? LENGTH: 6865  
 ? TYPE: DNA  
 ? ORGANISM: Homo sapiens  
 US-09-854-856 49

Query March 77.8% Score 147 DB 47 Length 6865  
 Best Local Similarity 100.0% Pred. E-value 0.0  
 Matches 147 Conservative 0 Mismatches 0 Indels 0 Gaps 0

QY 1 TCACGTTGCTTCA 14  
 DB 2357 TCACGTTGCTTCA 2370

RESULT 14  
 US-09-854-856 33  
 ? Sequence 33, Application US/09854856  
 ? Patent No. 654252  
 ? GENERAL INFORMATION:  
 ? APPLICANT: Walke, D. Wade  
 ? APPLICANT: Hilburn, Erin  
 ? APPLICANT: Boroko, Gregory  
 ? APPLICANT: Turner, C. Alexander Jr.  
 ? TITLE OF INVENTION: No. 654252el Human Kinases and Polynucleotides  
 ? TITLE OF INVENTION: Encoding the Same  
 ? FILE REFERENCE: LEX-0178-USA  
 ? CURRENT APPLICATION NUMBER: US/09/854,856  
 ? PRIOR FILING DATE: 2001-05-14  
 ? PRIOR APPLICATION NUMBER: US 62/206,015  
 ? PRIOR FILING DATE: 2000-05-19  
 ? NUMBER OF SEQ ID NOS: 64  
 ? SOFTWARE: FastSeq for Windows Version 4.0  
 ? SEQ ID NO 33  
 ? LENGTH: 6969  
 ? TYPE: DNA  
 ? ORGANISM: Homo sapiens  
 US-09-854-856 33

Query March 77.8% Score 147 DB 47 Length 6969  
 Best Local Similarity 100.0% Pred. E-value 0.0  
 Matches 147 Conservative 0 Mismatches 0 Indels 0 Gaps 0

QY 1 TCACGTTGCTTCA 14  
 DB 2441 TCACGTTGCTTCA 2454

Search completed: October 21, 2003, 05:19:27  
 Job time : 4.41415 secs

Result No.	Score	Query Match	Length	DB ID	Description
1	18	100.0	216	22	ABR46557
2	18	100.0	216	22	ABR44410
3	18	100.0	216	22	AAK12378
4	18	100.0	216	22	AAK39607
5	18	100.0	216	22	AAI13407
6	18	100.0	216	22	AAI44595
7	18	100.0	216	22	AAI35112
8	18	100.0	216	23	AB538171





CC epilepsy and cancers. The present sequence is one of the probes of the  
 CC invention.

XX Sequence 216 BP; 39 A; 55 C; 60 G; 62 T; 0 other;

Query Match 100.0%; Score 18; DB 22; Length 216;  
 Best Local Similarity 100.0%; Pred. No. 8.7;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY : TCACGTTGGCTTCAGAGG 18  
 : |||||  
 Db : TCACGTTGGCTTCAGAGG 18

## RESULT 4

AAK38607  
 ID AAK38607 standard; DNA; 216 BP.

XX

AC AAK38607;

XX 06-NOV-2001 (first entry)

XX Human bone marrow expressed single exon probe SEQ ID NO: 13164.

DE Human bone marrow expressed single exon probe SEQ ID NO: 13164.  
 XX Human bone marrow expressed exon; gene expression analysis; probe;  
 KW microarray; cancer; leukemia; lymphoma; myeloma; ss.

XX Homo sapiens.

OS

XX

XX WQ200157276-A2.

XX

XX 09-AUG-2001.

XX

XX 30-JAN-2001; 2001WC-US00569.

XX

XX 24-FEB-2000; 2000US-0180312.

XX

XX 26-MAY-2000; 2000US-0207456.

XX

XX 30-JUN-2000; 2000US-0608408.

XX

XX 03-AUG-2000; 2000US-0632366.

XX

XX 21-SEP-2000; 2000US-0234487.

XX

XX 27-SEP-2000; 2000US-0234359.

XX

XX 24-OCT-2000; 2000GB-0024263.

XX

PA (NCBI-) MOLECULAR DYNAMICS INC.

XX

XX Form SG, Hanzel DK, Chen W, Rank DR;

XX

XX WPI; 2001-489901/533.

XX

XX Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human bone marrow

XX

XX Example 4; SEQ ID NO: 13164; 688bp \* Sequence listing English.

XX

XX The present invention provides a number of single exon nucleic acid

CC probes which are derived from genomic sequences expressed in the human

CC bone marrow. They can be used to measure gene expression in bone marrow

CC samples, which may enable the improved diagnosis and treatment of cancers

CC such as lymphoma, leukemia and myeloma. The present sequence is one of

CC the probes of the invention.

XX

XX Sequence 216 BP; 39 A; 55 C; 60 G; 62 T; 0 other;

SC

Query Match 100.0%; Score 18; DB 22; Length 216;

Best Local Similarity 100.0%; Pred. No. 8.7;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY : TCACGTTGGCTTCAGAGG 18

: |||||

Db : TCACGTTGGCTTCAGAGG 18

## RESULT 5

AA119407

ID AA119407 standard; DNA; 216 BP.

XX

AC AA119407;

XX

XX 12-OCT-2001 (first entry)

XX

XX Probe #3140 for gene expression analysis in human cervical cell sample.

DE

XX Probe; human; microarray; gene expression; cervical epithelial cell;

XX cervical cancer; ss.

XX

OS Homo sapiens.

XX

XX WQ200157276-A2.

XX

XX 29-AUG-2001.

XX

XX 30-JAN-2001; 2001WC-US00569.

XX

XX 04-FEB-2000; 2000US-0180312.

XX

XX 26-MAY-2000; 2000US-0207456.

XX

XX 30-JUN-2000; 2000US-0608408.

XX

XX 21-AUG-2000; 2000US-0632366.

XX

XX 21-SEP-2000; 2000US-0234487.

XX

XX 27-SEP-2000; 2000US-0234359.

XX

XX 24-OCT-2000; 2000GB-0024263.

XX

PA (NCBI-) MOLECULAR DYNAMICS INC.

XX

XX Form SG, Hanzel DK, Chen W, Rank DR;

XX

XX WPI; 2001-489901/533.

XX

XX Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human cervical epithelial cells

XX

XX Claim 25; SEQ ID No 9140; 487bp; English.

XX

XX The present invention relates to human single exon nucleic acid probes

CC (SENPs). The present sequence is one such probe. The SENPs are derived

CC from human HeLa cells. The SENPs can be used to produce a single exon

CC microarray, which can be used for measuring human gene expression in a

CC sample derived from human cervical epithelial cells. By measuring gene

CC expression, the probes are therefore useful in grading and/or staging

CC of diseases of the cervix, notably cervical cancer.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from W.P.C.

XX at [http://wipo.int/pub/published\\_pat\\_sequences](http://wipo.int/pub/published_pat_sequences).

XX

XX Sequence 216 BP; 39 A; 55 C; 60 G; 62 T; 0 other;

SC

Query Match 100.0%; Score 18; DB 22; Length 216;

Best Local Similarity 100.0%; Pred. No. 8.7;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY : TCACGTTGGCTTCAGAGG 18

: |||||

Db : TCACGTTGGCTTCAGAGG 18

## RESULT 6

AA144595

ID AA144595 standard; DNA; 216 BP.

XX

AC AA144595;

XX

XX 17-OCT-2001 (first entry)

XX

XX Probe #13281 used to measure gene expression in human placenta sample.

DE

XX Probe; microarray; human; placenta; antenatal diagnosis;

XX genetic disorder; ss.

KW

```

XX OS Homo sapiens.
XX PN WO200:57272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001:WO-US00663.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0606428.
XX PR 03-AUG-2000; 2000US-0642366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX P: Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-498898/53.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta.
XX PS Claim 25; SEQ ID No 13281; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SENPs).
XX CC The present sequence is one such probe. The probes are useful for
XX CC producing a microarray for predicting, measuring and displaying gene
XX CC expression in samples derived from human placenta. The probes are useful
XX CC for antenatal diagnosis of human genetic disorders.
XX SQ Sequence 216 BP; 39 A; 55 C; 60 G; 62 T; 0 other;
    Query Match 100.0%; Score 18; DB 22; Length 216;
    Best Local Similarity 100.0%; Pred. No. 8.7;
    Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
    Oy : TCACGTTGGCTTCACAGG 18
    Db 1 TCACGTTGGCTTCACAGG 18
RESULT 7
AA105132
ID AA105132 standard; DNA; 216 BP.
XX AC AA105132;
XX DT 09-OCT-2001 (first entry);
XX DE Probe #5123 used to measure gene expression in human breast sample.
XX KW Probe; human; breast disease; breast cancer; development disorder; ss;
XX KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX OS Homo sapiens.
XX PN WO200:57272-A2.
XX PD 09-AUG-2001.
XX PF 29-JAN-2001; 2001:WO-US00663.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0606408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.

```

```

XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX P: Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-476286/51.
XX PT Novel single exon nucleic acid probe used to measuring gene expression
XX in a human breast
XX PS Claim 25; SEQ ID No 5123; 122pp; English.
XX CC The present invention relates to novel single exon nucleic acid probes.
XX CC The present sequence is one such probe. The probes are useful for
XX CC measuring human gene expression in a human breast sample, where the probe
XX CC hybridises at high stringency to a nucleic acid expressed in the human
XX CC breast. The probes are useful for predicting, diagnosing, grading,
XX CC staging, monitoring and premeasuring diseases of the human breast,
XX CC particularly those diseases with polyploid activity. The diseases
XX CC include breast cancer, the various development, inflammatory diseases
XX CC of the breast, fibrocystic conditions, proliferative breast disease and
XX CC non-carcinoma tumours.
XX CC Note: The sequence data for this patent did not form part of the claimed
XX CC specification, but was obtained in electronic format directly from WPI.
XX CC at ftp.wpi.com/pub/published-pat-sequences.
XX SQ Sequence 216 BP; 39 A; 55 C; 60 G; 62 T; 0 other;
    Query Match 100.0%; Score 18; DB 22; Length 216;
    Best Local Similarity 100.0%; Pred. No. 8.7;
    Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
    Oy : TCACGTTGGCTTCACAGG 18
    Db 1 TCACGTTGGCTTCACAGG 18
RESULT 8
ABS31671
ID ABS31671 standard; DNA; 216 BP.
XX AC ABS31671;
XX DT 25-FEB-2003 (first entry);
XX DE Human liver single exon probe, SEQ ID No 13161.
XX KW Human; single exon nucleic acid probe; liver; cirrhosis;
XX KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
XX KW coronary heart disease; ss.
XX OS Homo sapiens.
XX PN WO200157273 A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001:WO-US01564.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0606428.
XX PR 03-AUG-2000; 2000US-0642366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX P: Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-498898/53.

```

```

PT Human genome-derived single exon nucleic acid probes useful for
PT analysing gene expression in human adult liver
XX
XX
XX Claim 4: SEQ ID No 13161; 634pp; English.
XX
XX The invention relates to a single exon nucleic acid probe (SENPI) (i) for
XX measuring human gene expression in a sample derived from human adult
XX liver, comprising one of 13109 defined nucleotide sequences given in the
XX specification (or complements/fragments). The probe hybridises at high
XX stringency to a nucleic acid molecule expressed in the human adult
XX liver. (ii) may be used for predicting, measuring and displaying gene
XX expression in samples derived from human adult liver. The genes
XX identified may be involved in genetic liver diseases such as cirrhosis,
XX hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which
XX is associated with coronary heart disease. AAS25011 AAS51005 represent
XX human liver single exon nucleic acid probes of the invention.
XX Note: The sequence information for this patent does not appear in the
XX printed specification but was obtained in electronic format directly
XX from WFO at ftp.wipo.int/pub/published.pat_sequences.
XX
XX Sequence 216 BP; 39 A; 55 C; 60 G; 62 T; 0 other;
XX
XX Query Match 100.0%; Score 19; DB 24; Length 216;
XX Best Local Similarity 100.0%; Pred. No. 8.7;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 TCACGTTGGCTTCAGAG 19
XX DB 1 TCACGTTGGCTTCAGAG 19
XX
XX RESULT 9
XX ID AAS12673 standard; DNA; 216 BP.
XX AC AAS12673;
XX
XX 13 AUG 2002 (first entry)
XX
XX DE Human genome-derived single exon probe QSP from the SEQ ID No 12664.
XX
XX Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
XX chronic obstructive pulmonary disease; interstitial lung disease;
XX familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX Niemann-Pick disease; Hermansky-Pudlak syndrome; sarcoidosis; pulmonary
XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
XX primary ciliary dyskinesia; pulmonary hypertension;
XX hyaline membrane disease; open reading frame; (ds)
XX
XX Homo sapiens.
XX
XX OS Homo sapiens.
XX
XX PK WC200186903-A2.
XX
XX 15-NOV-2001.
XX
XX 30-JAN-2001; 2001WC-US000665.
XX
XX 04-FEB-2000; 2000US-180312P.
XX 26-MAY-2000; 2000US-207456P.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0612366.
XX 21-SEP-2000; 2000US-234687P.
XX 27-SEP-2000; 2000US-236352P.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WP; 2002-114193/15.
XX

```

```

PT Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples -
XX
XX Claim 4: SEQ ID No 12664; 634pp; English.
XX
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human lung comprising single exon nucleic acid probes having one of
XX 12614 nucleic acid sequences mentioned in the specification, or their
XX complements or the 12617 open reading frames derived from the 12614
XX probes. Also included are a microarray comprising the novel set of
XX probes. The novel set of probes which hybridise at high stringency to a
XX sample derived from human lung, comprising (a) contacting the array with
XX a collection of detectably labeled nucleic acids derived from human lung
XX mRNA, and (b) measuring the label detectably bound to each probe of
XX the array; identifying exons in a eukaryotic genome, comprising
XX (i) algorithmically predicting at least one exon from genomic sequences
XX of the eukaryote; and (b) detecting specific hybridisation of detectably
XX labeled nucleic acids from eukaryotic lung mRNA, to a single exon probe,
XX having a fragment identical to the predicted exon, the probe is included
XX in the above mentioned microarray; assigning exons to a single gene.
XX comprising (a) identifying exons from genomic sequence by the method
XX above and (b) measuring the expression of each of the exons in several
XX tissues and/or cell types using hybridisation to a single exon
XX microarrays having a probe with the exon, where a common pattern of
XX expression of the exons in the tissues and/or cell types indicates that
XX the exons should be assigned to a single gene; a peptide comprising one
XX of 12011 sequences, mentioned in the specification, or encoded by the
XX probes/open reading frames (OSF). The probes are used for gene
XX expression analysis, and for identifying exons in a gene, particularly
XX using human lung derived mRNA and for the study of lung diseases
XX such as asthma, lung cancer, chronic obstructive pulmonary disease
XX (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
XX fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
XX Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
XX haemorrhoidosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
XX pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
XX pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
XX and hyaline membrane disease. The present sequence is a single exon
XX probe open reading frame of the invention.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WFO at
XX ftp.wipo.int/pub/published.pat_sequences.
XX
XX Sequence 216 BP; 39 A; 55 C; 60 G; 62 T; 0 other;
XX
XX Query Match 100.0%; Score 19; DB 24; Length 216;
XX Best Local Similarity 100.0%; Pred. No. 8.7;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 TCACGTTGGCTTCAGAG 19
XX DB 1 TCACGTTGGCTTCAGAG 19
XX
XX RESULT 10
XX AAS1741
XX ID AAS1741 standard; DNA; 478 BP.
XX AC AAS1741;
XX
XX 01-FEB-2002 (first entry)
XX
XX DE Human foetal liver single exon nucleic acid probe #46.
XX
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
XX Homo sapiens.
XX
XX OS Homo sapiens.
XX
XX PK WC200157277-A2.
XX

```



CC Bone marrow. They can be used to measure gene expression in bone marrow  
 CC samples, which may enable the improved diagnosis and treatment of cancers  
 CC such as lymphoma, leukemia and myeloid. The present sequence is one of  
 CC the probes of the invention.

XX SQ Sequence 478 BP; 136 A; 131 G; 100 G; 131 T; 0 Other;

Query Match 100.0%; Score 18; DB 22; Length 478;  
 Best Local Similarity 100.0%; Pred. No. 9.3;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 TCACGTTGGCTTCAGAG 18  
 Dc 270 TCACGTTGGCTTCAGAG 287

# RESULT 15

AA10052  
 ID AA10052 standard; DNA; 478 BP;  
 AC AA10052;  
 XX  
 XX  
 XX

DT 12-OCT-2001 (first entry)

DE Probe #43 for gene expression analysis in human cervical cell sample.

XX Probe; human; microarray; gene expression; cervical epithelial cell;  
 XX cervical cancer; ss.

XX Homo sapiens.

XX WO/01/57272-A2.

XX 09 AUG 2001.

XX 30 JAN 2001; 2001WO-US00663.

XX 04 FEB 2000; 2000US-0190312.

XX 26 MAY 2000; 2000US-0207456.

XX 30 JUN 2000; 2000US-0608408.

XX 03 AUG 2000; 2000US-0632366.

XX 21 SEP 2000; 2000US-0244687.

XX 27 SEP 2000; 2000US-0236359.

XX 04 OCT 2000; 2000GB-C024263.

XX (MOLECULAR DYNAMICS INC.)

XX Penn SS, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488897/53.

XX Human genome derived single exon nuclear acid probes useful for  
 XX analyzing gene expression in human cervical epithelial cells.

XX Claim 25; SEQ ID No 44; 487bp; English.

XX The present invention relates to human single exon nuclear acid probes  
 CC (SENPs). The present sequence is one such probe. The SENPs are derived  
 CC from human HeLa cells. The SENPs can be used to provide a single exon  
 CC microarray, which can be used for measuring human gene expression in a  
 CC sample derived from human cervical epithelial cells by measuring gene  
 CC expression. The probes are therefore useful in grading and/or staging  
 CC of diseases of the cervix, notably cervical cancer.

CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WFO  
 CC at ftp.wipo.int/pub/published\_pat\_sequences.

XX Sequence 478 BP; 136 A; 131 G; 100 G; 131 T; 0 Other;

Query Match 100.0%; Score 18; DB 22; Length 478;  
 Best Local Similarity 100.0%; Pred. No. 9.3;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 TCACGTTGGCTTCAGAG 18  
 Dc 270 TCACGTTGGCTTCAGAG 287

# RESULT 14

AA131357  
 ID AA131357 standard; DNA; 478 BP;  
 AC AA131357;  
 XX  
 XX  
 XX

DT 17-OCT-2001 (first entry)

DE Probe #43 used to measure gene expression in human placenta sample.

XX Probe; microarray; human; placenta; antenatal diagnosis;  
 XX genetic disorder; ss.

XX Homo sapiens.

XX WO200157272-A2.

XX 09 AUG 2001.

XX 30 JAN 2001; 2001WO-US00663.

XX 04 FEB 2000; 2000US-0190312.

XX 26 MAY 2000; 2000US-0207456.

XX 30 JUN 2000; 2000US-0608408.

XX 03 AUG 2000; 2000US-0632366.

XX 21 SEP 2000; 2000US-0244687.

XX 27 SEP 2000; 2000US-0236359.

XX 04 OCT 2000; 2000GB-C024263.

XX (MOLECULAR DYNAMICS INC.)

XX Penn SS, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488897/53.

XX Human genome-derived single exon nuclear acid probes useful for  
 XX analyzing gene expression in human placenta.

XX Claim 25; SEQ ID No 43; 654bp; English.

XX The present invention relates to single exon nuclear acid probes (SENPs).  
 CC The present sequence is one such probe. The probes are useful for  
 CC predicting a microarray for predicting, measuring and displaying gene  
 CC expression in samples derived from human placenta. The probes are useful  
 CC for antenatal diagnosis of human genetic disorders.

XX Query Match 100.0%; Score 18; DB 22; Length 478;

XX Best Local Similarity 100.0%; Pred. No. 9.3;

XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 TCACGTTGGCTTCAGAG 18  
 Dc 270 TCACGTTGGCTTCAGAG 287

# RESULT 15

AA10052  
 ID AA10052 standard; DNA; 478 BP;  
 AC AA10052;  
 XX  
 XX  
 XX

DT 09-OCT-2001 (first entry)

DE Probe #43 used to measure gene expression in human breast sample.

XX Probe; human; breast disease; breast cancer; development disorder; ss;

```

KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX Homo sapiens.
OS
XX WO2001:57270-A2.
XX
XX 09 AUG 2001.
XX
XX 29-JAN-2001; 2001:WO-US00661.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY 2000; 2000US-0207456.
XX 30-JUN 2000; 2000US-0608458.
XX 03-AUG-2000; 2000US-0612366.
XX 21-SEP-2000; 2000US-0214687.
XX 27-SEP-2000; 2000US-0216359.
XX 04-OCT-2000; 2000GB-0242631.
XX
XX (MOLE ) MOLECULAR DYNAMICS INC.
XX Penn SQ, Hanzei DX, Chen W, Rank DR.
XX WPI; 2001-476255/51.
XX
XX Novel single exon nucleic acid probe used to measuring gene expression
XX in a human breast .
XX
XX Claim 95; SEQ ID No 43; 322pp; English.
XX
XX The present invention relates to novel single exon nucleic acid probes.
XX The present sequence is one such probe. The probes are useful for
XX measuring human gene expression in a human breast sample, where the probe
XX hybridises at high stringency to a nucleic acid expressed in the human
XX breast. The probes are useful for predicting, diagnosing, grading,
XX staging, monitoring and prognosing diseases of the human breast,
XX particularly those diseases with polygenic aetiology. The diseases
XX include breast cancer, disorders of development, inflammatory diseases
XX of the breast, fibrocystic changes, proliferative breast disease and
XX non-carcinoma tumours.
XX
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pat_sequences.
XX
XX Sequence 478 BP; 136 A; 111 C; 100 G; 131 T; 0 other;
XX
XX Query Match 100.0%; Score 18; DH 22; Length 478;
XX Residual Similarity 100.0%; Pred. No. 9.3;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 TCACGTTGGCTTCAGAGG 18
XX ||| |||| ||||
XX 270 TCACGTTGGCTTCAGAGG 287
XX
XX Search completed: October 21, 2003, 00:03:03
XX Job time : 12.3994 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Coriagen Ltd.

ON nucleic nucleic search, using sw model

Run on: October 20, 2003, 18:30:49 : Search time 180.456 Seconds  
(without alignments)  
4354.525 Million cell updates/sec

File: us-10-029-137-7

Perfect score: 15

Sequence: 1 aaacgttgagcttcagagg 18

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 2278:392 seqs, 1215223656 residues

Total number of hits satisfying chosen parameters: 455e2784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST\*

1: em\_estb1\*

2: em\_esth1\*

3: em\_estm1\*

4: em\_estm1\*

5: em\_estov1\*

6: em\_estp1\*

7: em\_estp1\*

8: em\_h1c1\*

9: gb\_est1\*

10: gb\_est2\*

11: gb\_h1c1\*

12: gb\_est1\*

13: gb\_est1\*

14: gb\_est1\*

15: em\_estfun\*

16: em\_estm1\*

17: em\_gss\_h1c1\*

18: em\_gss\_h1c1\*

19: em\_gss\_h1c1\*

20: em\_gss\_v1\*

21: em\_gss\_h1c1\*

22: em\_gss\_h1c1\*

23: em\_gss\_h1c1\*

24: em\_gss\_h1c1\*

25: em\_gss\_h1c1\*

26: em\_gss\_h1c1\*

27: em\_gss\_h1c1\*

28: gb\_gss1\*

29: gb\_gss2\*

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15	100.0	251	9	A1539573
2	15	100.0	276	12	BQ011381
3	18	100.0	279	12	BQ013938
4	18	100.0	365	9	A1538977

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

C	5	18	100.0	382	14	CA338172
	6	18	100.0	403	14	W88618
C	7	18	100.0	410	14	W93361
	8	18	100.0	417	9	AA039643
	9	18	100.0	425	9	AA030019
	10	18	100.0	437	9	AA012894
	11	18	100.0	437	9	AA012894
	12	18	100.0	444	9	AA012894
	13	18	100.0	456	9	AA012894
	14	18	100.0	461	9	AA012894
	15	18	100.0	465	9	AA012894
C	16	18	100.0	465	14	CA218142
	17	18	100.0	465	14	H97895
	18	18	100.0	469	9	AA012894
	19	18	100.0	477	9	AA012894
C	20	18	100.0	480	14	W93361
	21	18	100.0	487	13	BX282583
C	22	18	100.0	489	14	W67773
	23	18	100.0	500	9	AA012894
C	24	18	100.0	526	13	BC071817
	25	18	100.0	532	9	AA012894
	26	18	100.0	536	9	AA012894
	27	18	100.0	537	9	AA012894
	28	18	100.0	542	9	AA012894
	29	18	100.0	549	10	BF438680
C	30	18	100.0	557	12	BQ011000
	31	18	100.0	559	9	AA012894
C	32	18	100.0	569	9	AA012894
	33	18	100.0	569	10	BF438680
	34	18	100.0	569	10	BF438680
C	35	18	100.0	569	12	BQ011000
	36	18	100.0	569	12	BQ011000
C	37	18	100.0	570	10	BC011320
	38	18	100.0	573	14	CA952729
C	39	18	100.0	573	14	CA952729
	40	18	100.0	589	14	CA952729
	41	18	100.0	591	12	BQ011000
C	42	18	100.0	593	12	BQ011000
	43	18	100.0	595	12	BQ011000
	44	18	100.0	517	12	BQ011000
	45	18	100.0	521	14	CA952484

#### ALIGNMENTS

A1539573 251 bp mRNA linear EST 13 MAY 1999  
tp35b10.x: NCI-CGAP-Ur4 Homo sapiens cDNA clone IMAGE:2189755.3  
similar to contains A15 repetitive element; contains TARI.1; TARI  
TARI repetitive element; mRNA sequence.

A1539573 1 G:4453709

EST

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 251)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

Tumor Gene Index

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-raman@nih.gov](mailto:cgapbs-raman@nih.gov)

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emerit-Buck, M.D., Ph.D.

CGAP Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.N.A.G.E. Consortium/LLNL at:



www-bio.llnl.gov/bbrp/image/image.html  
 Insert Length: 683 Std Error: 0.20  
 Seq Primer: -4002 from G:bc0  
 High quality sequence stop: 203  
 POLYA=No.

#### FEATURES

source

##### Location/Qualifiers

```
1..251
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cis="IMAGE:2189755"
/tissue_type="serous papillary carcinoma, high grade, 2
pooled tumors"
/lab_host="DH1CB"
/clone_lib="NCI CGAP Ut4"
/notes="Organ: uterus; Vector: pCWV-SF06T6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: oligo 47;
Average insert size: 1.45 kb; Life Technologies catalog #:
11542-016"
```

46 a 71 c 70 g 64 t

#### BASE COUNT

ORIGIN

Query Match 100.0% Score 18; PS 9; length 251;

Best Local Similarity 100.0% Pred No 91;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACCTTGGCTTCAGAG 18

||||| ||||| |||||

74 TCACCTTGGCTTCAGAG 91

#### RESULT 2

BQ011981

LOCUS

DEFINITION U1-1-BC1p atg-a-06-0-U1.s: NCI CGAP P13 Homo sapiens cDNA clone

ACCESSION BQ011981

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

276 bp mRNA linear EST 26-MAR-2002  
 U1-1-BC1p atg-a-06-0-U1.s: NCI CGAP P13 Homo sapiens cDNA clone  
 U1-1-BC1p atg-a-06-0-U1.3', mRNA sequence.  
 BQ011981  
 EST: 1 bases 1 to 276  
 Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 NC1-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs@mail.nih.gov  
 Tissue Procurement: Dr. Steven Brown  
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Clone distribution information can be obtained  
 from Dr. M. Bento Soares, bento-soaresuiowa.edu  
 The following repetitive elements were found in this cDNA  
 sequence: 1-25, >AT rich#Low\_complexity  
 Seq primer: M13 FORWARD  
 POLYA=yes.

#### FEATURES

source

##### Location/Qualifiers

```
1..276
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_lib="NCI CGAP P13"
/notes="Organ: placenta; Vector: p7T3-Pac (Pharmacia) with
a modified polylinker; Site 1: EcoR I; Site 2: Not I;
NC1-CGAP P13 is a subtracted cDNA library constructed
```

according to Honaldis, Lennon and Soares, Genome Research, 6:791-804, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p7T3-Pac vector. The oligonucleotide used to prime the synthesis of first strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are GA, AGGAA. For additional information, contact: Bento Soares, bento-soaresuiowa.edu

TAG: 1-BC1p

TAG: Tissue: placenta; human 6 week

TAG: SEC: NA

BASE COUNT 68 a 67 g 81 t 1 others

#### ORIGIN

Query Match 100.0% Score 18; PS 12; length 276;

Best Local Similarity 100.0% Pred No 91;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACCTTGGCTTCAGAG 18

||||| ||||| |||||

74 TCACCTTGGCTTCAGAG 100

#### RESULT 3

BQ011981

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

273 bp mRNA linear EST 26-MAR-2002  
 U1-1-BC1p atg-a-06-0-U1.s: NCI CGAP P13 Homo sapiens cDNA clone  
 U1-1-BC1p atg-a-06-0-U1.3', mRNA sequence.  
 BQ011981  
 EST: 1 bases 1 to 273  
 Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 NC1-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs@mail.nih.gov  
 Tissue Procurement: Dr. Steven Brown  
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Clone distribution information can be obtained  
 from Dr. M. Bento Soares, bento-soaresuiowa.edu  
 The following repetitive elements were found in this cDNA  
 sequence: 1-24, >AT rich#Low\_complexity  
 Seq primer: M13 FORWARD  
 POLYA=yes.

#### FEATURES

source

##### Location/Qualifiers

```
1..279
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_lib="NCI CGAP P13"
/notes="Organ: placenta; Vector: p7T3-Pac (Pharmacia) with
a modified polylinker; Site 1: EcoR I; Site 2: Not I;
NC1-CGAP P13 is a subtracted cDNA library constructed
according to Bonaldi, Lennon and Soares, Genome Research,
6:791-804, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into p7T3-Pac
```





Washington University School of Medicine  
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: estawatson@wustl.edu

This clone is available royalty free through LNL; contact the  
IMAGE Consortium (<http://image.llnl.gov>) for further information.

Seq primer: -28M: rev2 from Amerham

High quality sequence stop: 315

#### FEATURES

##### SOURCE

1: 417  
/organism "Homo sapiens"  
/mol\_type="cDNA"  
/db\_xref="XREF:1384871"  
/db\_xref="taxon:9606"  
/clone="IMAGE:376315"  
/sex="unknown"  
/dev\_stage="19 weeks"  
/lab\_host="Chick (ampicillin resistant)"  
/cclone\_lib="Soares fetal heart cDNA library"  
/notes="Organ: heart; Vector: pTZ19 (Pharmacia) with a  
modified polylinker; Site 1: Not 1; Site 2: Eco RI; 1st  
strand cDNA was primed with a Not 1 - oligo(dT) primer (5'  
TGTTACCATCTGACAGTGGAGGCGCCATCTTTTCTTTTCTTTT 3'),  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not 1 and cloned into  
the Not 1 and Eco RI sites of a modified pTZ19 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by  
M. Fatima Bonaldo. This library was constructed from the  
same fetuses as the fetal lung library. Soares fetal lung  
cDNA library."

BASE COUNT 102 a 85 c 88 g 141 t 1 others

##### ORIGIN

Query Match 100.0%; Score 18; DB 9; Length 417;  
Best Local Similarity 100.0%; Pred No. 1.2e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACGTTGGGTTTCAGAG 18

||||| ||||| ||||| |||||

Db 102 TCACGTTGGGTTTCAGAG 85

##### RESULT 9

AA300019

LOCUS EST12434 Virus titer 1 Homo sapiens cDNA clone: mRNA sequence. EST 16-APR 1997

DEFINITION AA300019

ACCESSION AA300019

VERSION AA300019.1 GI:1942149

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 425)

Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Feldner, R.A., Bult

, C.J., Lee, N.H., McKress, B.F., Weinstock, K.B., Gocayne, J.D., White

, O., Sutton, G., Blake, J.A., Brandon, R.C., McMillan, C., Clayton, R.A.,

Cline, T.R., Cotton, M.D., Eickholt, J., Fink, D.D., Fitzgerald

, L.M., Fitzhugh, W.M., Fritchman, J., Gocke, J., Gocke, A.,

Gnehm, C.L., Hanra, M.C., Hedblom, E., Hickie, F.S., Kelley, J.M.,

Kelley, J.C., Liu, L.I., Marra, S.M., Metzker, J.M.,

Moreno, Palanques, R.F., McDonald, J.A., Nguyen, B.T., Pellegrino, S.M.,

Phillips, C.A., Ryder, S.E., Scott, T.L., Sack, D.M., Shirley, R.,

Small, K.V., Spriggs, T.A., Utterback, T.P., Winkler, J.F., Zeng,

Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,

Dimke, D., Feng, D.-F., Ferris, A., Fischer, C., Hastings, G.A., He, W.W.,

Hu, J.S., Greene, J.M., Gruber, A., Hudson, P., Kim, A.K., Kozak, D.,

Kirsch, C., Kung'u, G., Li, H., Weissner, P.S., Olson, H., Raymond, L.,

Wei, Y.F., King, J., Xu, C., Ye, G., Ruben, S.M., Dillon, E.J., Fannon

, M.R., Rosen, C.A., Hasekline, A.A., Fields, C., Fraser, C.M., and

Went, J.C.

#### TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

Other ESTs: EST12622 TH0174495

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Tel: 301-861-9050

Fax: 301-861-9043

E-mail: [kerlavage@igrr.org](mailto:kerlavage@igrr.org)

For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene

Index (<http://www.tigr.org/tdb/hgi/hgi.html>;

Seq primer: M31-21

#### FEATURES

##### SOURCE

1: 425

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="XREF:192954"

/db\_xref="taxon:9606"

/dev\_stage="adult"

/clone\_lib="Uterus tumor 1"

/notes="Organ: uterus; Vector: pBluescript SK-; Site 1:

Base Count 126 a 95 c 90 g 112 t 2 others

##### ORIGIN

Query Match 100.0%; Score 18; DB 9; Length 425;

Best Local Similarity 100.0%; Pred No. 1.2e+02;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACGTTGGGTTTCAGAG 18

||||| ||||| ||||| |||||

Db 275 TCACGTTGGGTTTCAGAG 293

##### RESULT 10

AA812894

LOCUS a-812894.s1 Scores testis NBT Homo sapiens cDNA clone: EST 31-DEC 1998

DEFINITION -RNA sequence.

ACCESSION AA812894.1 GI:2882358

VERSION AA812894

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 437)

NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: [cgaps-f@mail.nih.gov](mailto:cgaps-f@mail.nih.gov)

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo

, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

cDNA sequencing by: Washington University Genome Sequencing Center

clone distribution: NCI-CCAP clone distribution information can be

found through the T.M.A.G.E. Consortium/LLNL at:

[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Insert length: 1527 Std Error: 0.00

Seq primer: -40m13 fwd. ET from A-ersham

High quality sequence stop: 427.

Location/Qualifiers

1: 437

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

```

/cldone="1377245"
/sex="male"
/lab host="DH10B"
/clone.lib="Soares testis NHT"
/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer; 15'
TGTTACCAATCTGAAGTGGAGCGCGCCCAATTTTCTTTTCTTTT 3'
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p773 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      130 a      97 c      94 g      116 t
ORIGIN
Query Match      100.0%; Score 18; DB 9; Length 437;
Best Local Similarity 100.0%; Pred. NO. 1.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY      1 TCACGTTGCGTTCAGAGC 18
      ||||| ||||| |||
DB      280 TCACGTTGCGTTCAGAGC 297

RESULT 11
AA431715      437 bp      mRNA      linear      EST 22-MAY 1997
LOCUS      zw77ec7.s1 Soares testis_NHT Homo sapiens cDNA clone IMAGE:782244
DEFINITION      3' mRNA sequence.
ACCESSION      AA431715
VERSION      AA431715.1 GI:2115423
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 437)
Kucaba, T., Lacy, M., Le, N., Lemmon, G., Maria, M., Martin, J., Moore, B.,
Schillerberg, X., Stogton, M., Tan, P., Theisinger, R., White, Y., Wylie
, T., Waterston, R., and Wilsch, R.
WashU-Merck EST Project 1997
Unpublished
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: es@watson.wustl.edu
This clone is available royalty free through INL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -41m3 fwd. ET from Amersham
High quality sequence stop: 396.
Location/Qualifiers
1..437
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:782244"
/sex="male"
/lab host="DH10B"
/clone.lib="Soares testis_NHT"
/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer; 15'
TGTTACCAATCTGAAGTGGAGCGCGCCCAATTTTCTTTTCTTTT 3'
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p773 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      130 a      97 c      94 g      116 t
ORIGIN
Query Match      100.0%; Score 18; DB 9; Length 437;
Best Local Similarity 100.0%; Pred. NO. 1.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY      1 TCACGTTGCGTTCAGAGC 18
      ||||| ||||| |||
DB      280 TCACGTTGCGTTCAGAGC 297

RESULT 12
AA741471      444 bp      mRNA      linear      EST 15-JAN-1996
LOCUS      z20.b29.s1 NCI NCI-60 Homo sapiens cDNA clone IMAGE:1286553
DEFINITION      NCI-60AP NCI-P1; www.ncbi.nlm.nih.gov/ncicgap.
ACCESSION      AA741471
VERSION      AA741471.1 GI:2115423
KEYWORDS      EST.
SOURCE      Homo sapiens
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 444)
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished
Contact: Robert Strauch, Ph.D.
Email: rstrauch@nci.nih.gov
Tissue Procurement: Lewis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Analyzed by: Greg Leeson, Ph.D.
cDNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-60AP clone distribution information can be
found through the NCI-60AP E. Consortium/INL at:
www.bioinformatics.nci.nih.gov/ncicgap.html
Seq primer: -40m3 fwd. ET from Amersham
High quality sequence stop: 416.
Location/Qualifiers
1..444
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1286553"
/sex="male"
/lab host="DH10B"
/clone.lib="NCI-60AP_G31"
/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD24+ IgD+),
prepared by Dr. Louis M. Staudt (NCI), Dr. David Allman,
(MNCI) and Dr. Gerald Marti (CHBR). cDNA synthesis was
performed with a Not I - oligo(dT) primer
15'-TGTTACCAATCTGAAGTGGAGCGCGCCCAATTTTCTTTTCTTTT-3'
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p773 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      141 a      91 c      87 g      125 t
ORIGIN
Query Match      100.0%; Score 18; DB 9; Length 444;
Best Local Similarity 100.0%; Pred. NO. 1.2e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY      1 TCACGTTGCGTTCAGAGC 18

```



Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www-bio.llnl.gov/bbrp/image.html](http://www-bio.llnl.gov/bbrp/image.html)  
 Insert Length: 784 Std Error: 0.00  
 Seq primer: 40UP from Gibco  
 High quality sequence: 423.  
 Location/Qualifiers

FEATURES

1..465  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE1917106"  
 /issue\_type="2 pooled tumors (clear cell type)"  
 /map\_pos="PH108"  
 /clone\_lib="NCI-CCAP Kid5"  
 /note="Organ: kidney; Vector: pMT73 Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I : oligo(dT) primer (5' AACTGGAGATTCGCCGCCGATTTTCTTTTCTTTT 3'), double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pMT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Ronaldo."

BASE COUNT 129 a 108 c 96 g 132 t  
 ORIGIN

Query Match 100.0%; Score 18; DP 9; Length 465;  
 Best local Similarity 100.0%; Pred.No. size+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TCACCTTCGCTTCAGG 18  
 |||||  
 Db 257 TCACCTTCGCTTCAGG 274

Search completed: October 21, 2003, 02:09:42  
 Job time : 106.466 secs





```

REFERENCE 2 (bases 1 to 789)
AUTHORS Mat.N. and Liu.C.
TITLE Direct Submission
JOURNAL Submitted 01 NOV-2000; Surgical Oncology, University of
California, Los Angeles, 10833 Le Conte Ave., Los Angeles, CA
90095, USA
FEATURES
Source Location/Qualifiers
1..789
/organism="Homo sapiens"
/mol_type="rRNA"
/db_xref="taxon:9606"
12..593
/codon_start=1
/product="tumor-related protein"
/db_xref="GI:13021850"
/translation="MAAPLGMPGCGPQPCAPPLPQASLLQAPAPPPSSSTL
VDELFSSFEAFASLVQVYNGTCEIRITGVQCCICQELDIARTPEFFLEFQIC
SVQKPEVQIKDQVLEKELQKDALVQKHLKQKPHQVLEINLVYKRPATPLPIS
LAYLQAGIQCPQCTSEANVSNGQQLAYEWAGA"
BASE COUNT 210 a 186 c 199 g 198 t 0 d n t s
ORIGIN
Query Match 100.0%; Score 18; DB 6; Length 789;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACGTTGGCTTCAGAGG 18
|||||
DB 549 TCACGTTGGCTTCAGAGG 532

RESULT 2
BC015361/c
LOCUS Human protein and cDNA[4].
DEFINITION BC015361
ACCESSION BC015361.1 GI:22556499
KEYWORDS LP 200218584-A/7.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Kato,S. and Saeki,M.
TITLE Human protein and cDNA[4].
JOURNAL Patent: JP 2001218584-A 7 14-AUG-2001;
JAPAN SCIENCE AND TECHNOLOGY CORP
COMMENT OS Homo sapiens (human)
PN JP 2001218584-A/7
PD 14-AUG-2001
PF 08-FEB-2000 JP 2000031062
PI SEIJI KATO,MHORO SAEKI
PC C12N15/09,C07K14/435,C07K14/43,C12N1/16,C12N1/19,C12N1/21, Po
C12N5/1077
CC C12P21/08,C12N15/09,C12N5/00
PH Key Location/Qualifiers
FT CDS (10)..(546)
FEATURES
Source Location/Qualifiers
1..822
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
210 a 186 c 199 g 227 t
BASE COUNT 210 a 186 c 199 g 227 t
ORIGIN
Query Match 100.0%; Score 18; DB 6; Length 822;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACGTTGGCTTCAGAGG 18
|||||

```

```

DB 546 TCACGTTGGCTTCAGAGG 543

RESULT 3
AX014817/c
LOCUS Sequence 3 from Homo sapiens
DEFINITION AX014817
ACCESSION AX014817
VERSION AX014817.1 GI:11741084
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 Schmitt,A., Spelsberg,T., Dubler,E., Huetzmann,R., Rosenthal,A. and
Pilatczyk,P.
AUTHORS Human protein and cDNA[4].
TITLE Human protein and cDNA[4].
JOURNAL Submitted 01 NOV-2000; Surgical Oncology, University of
California, Los Angeles, 10833 Le Conte Ave., Los Angeles, CA
90095, USA
KEYWORDS LP 200218584-A/7.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Kato,S. and Saeki,M.
TITLE Human protein and cDNA[4].
JOURNAL Patent: JP 2001218584-A 7 14-AUG-2001;
JAPAN SCIENCE AND TECHNOLOGY CORP
COMMENT OS Homo sapiens (human)
PN JP 2001218584-A/7
PD 14-AUG-2001
PF 08-FEB-2000 JP 2000031062
PI SEIJI KATO,MHORO SAEKI
PC C12N15/09,C07K14/435,C07K14/43,C12N1/16,C12N1/19,C12N1/21, Po
C12N5/1077
CC C12P21/08,C12N15/09,C12N5/00
PH Key Location/Qualifiers
FT CDS (10)..(546)
FEATURES
Source Location/Qualifiers
1..822
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
210 a 186 c 199 g 227 t
BASE COUNT 210 a 186 c 199 g 227 t
ORIGIN
Query Match 100.0%; Score 18; DB 6; Length 822;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACGTTGGCTTCAGAGG 18
|||||
DB 546 TCACGTTGGCTTCAGAGG 532

RESULT 4
AF137073/c
LOCUS Sequence 3 from Homo sapiens
DEFINITION AF137073
ACCESSION AF137073
VERSION AF137073.1 GI:11741084
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Mat.N. and Liu.C.
TITLE A gene from Homo sapiens
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 789)
AUTHORS Mat.N. and Liu.C.
TITLE Direct Submission
JOURNAL Submitted 01 NOV-2000; Surgical Oncology, University of
California, Los Angeles, 10833 Le Conte Ave., Los Angeles, CA
90095, USA
FEATURES
Source Location/Qualifiers
1..822
/organism="Homo sapiens"
/mol_type="rRNA"
/db_xref="taxon:9606"
12..593
/codon_start=1
/product="tumor-related protein"
/db_xref="GI:13021850"
/translation="MAAPLGMPGCGPQPCAPPLPQASLLQAPAPPPSSSTL
VDELFSSFEAFASLVQVYNGTCEIRITGVQCCICQELDIARTPEFFLEFQIC
SVQKPEVQIKDQVLEKELQKDALVQKHLKQKPHQVLEINLVYKRPATPLPIS
LAYLQAGIQCPQCTSEANVSNGQQLAYEWAGA"
BASE COUNT 210 a 186 c 199 g 198 t 0 d n t s
ORIGIN
Query Match 100.0%; Score 18; DB 6; Length 822;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACGTTGGCTTCAGAGG 18
|||||
DB 546 TCACGTTGGCTTCAGAGG 532

```

```

QY      1 TCACGTGGCTTCAGAGG 18
      1 ||||| ||||| |||||
Cb      541 TCACGTGGCTTCAGAGG 524

RESULT 5
AX017517/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Schmitt,A., Specht,T., Dahl,E., Hinzman,B., Rosenthal,A. and
Pilarczyk,C.
Human nucleic acid sequences from normal breast tissue
Patent: WO 9947655-A 65 23-SEP-1999;
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EUGEN (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILARSKY CHRISTIAN (DE)
Location/Qualifiers
1..1071
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT      281 a 218 c 255 g 317 t
ORIGIN
Query Match      100.0%; Score 18; DB 6; Length 1071;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TCACGTGGCTTCAGAGG 18
      1 ||||| ||||| |||||
Cb      546 TCACGTGGCTTCAGAGG 529

RESULT 7
BC011936/c
LOCUS
DEFINITION
Homo sapiens, hypothetical protein DK22p44N185, clone WGC22562
IMAGE:3936713, mRNA, complete cds.
ACCESSION
BC011936
VERSION
BC1:1936.1 GI:15090366
KEYWORDS
Homo sapiens (human)
SOURCE
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Strausberg,B.
Direct Submission
Submitted (30 JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
E-mail: mgc@pds-1.nhlbi.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
cDNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NIHSC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc@ncbigr.nih.gov
Shevchenko,Y., Wetherby,K.D., Beckstrom-Sternberg,S.M.,
Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brickley,C., Brooks,S.,
Dietrich,N.L., Guan,X., Gupta,J., Ho,S.-J., Karlins,E., Legaspi,A.,
Lim,M., Madroski,C.L., Mastello,C., Nairan,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Snyder,B., Stantrisp,S., Thomas,P.C.,
Tongson,E.E., Touchman,S.W., Tsugeon,C., Vogt,J.D., Walker,M.A.,
Zhang,J.-H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 28 Row: f Column: 4
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.
Location/Qualifiers
1..1083
/organism="Homo sapiens"
/mol_type="cDNA"
/db_xref="taxon:9606"
/clone="WGC22562 IMAGE:3936713"
/tissue_type="lung, small cell carcinoma"
FEATURES
source

```

```

/clone_lib="NH_MGC_7"
/adb_host="DH08-R"
/notes="Vector: pOT87"
14..550
/codon_start:
/product="hypothetical protein DKFp434N185"
/protein_id="AAH1936.1"
/db_xref="GI:15280367"
/translations="MAAPLGCMFSGGPPGPPQAPGLPQASLLQAAAGCAPRPSSTL
VQLESSEFACFASLVSDYVNGTDOEERTVDDCCKFLDIARQTECFKCKLQLOL
SVQKSEVIEDVSELRNLRKQCALVCKELKLRHKQCVLEUINVQHKRPADIQGS
LAYEQSAN:PAPEKPT"
314 t
BASE COUNT 294 a 216 c 259 g 314 t
ORIGIN
Query Match 100.0%; Score 18; DB 9; Length 1081;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCACGTTGGCTTCAGAG 14
+ ||||| |||||
Db 546 TCACGTTGGCTTCAGAG 533

RESULT 8
AX015350/c 1088 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 259 from Patent WO9553040.
ACCESSION AX015350
VERSION AX015350.1 GI:10041189
KEYWORDS Homo sapiens (human);
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Pilarczyk,C.
HUMAN nucleic acid sequences from ovarian tumour tissue
PATENT: WO 9553040-A 259 21-OCT-1999;
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILARSKY CHRISTIAN (DE);
LOCATION/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
302 a 216 c 254 g 316 t
BASE COUNT 302 a 216 c 254 g 316 t
ORIGIN
Query Match 100.0%; Score 18; DB 6; Length 1088;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCACGTTGGCTTCAGAG 14
+ ||||| |||||
Db 546 TCACGTTGGCTTCAGAG 529

RESULT 9
AF358829/c 1363 bp mRNA linear PRI 11-JAN-2002
LOCUS Homo sapiens tumor angiogenesis marker mRNA, complete cds.
DEFINITION Homo sapiens
ACCESSION AF358829
VERSION AF358829.1 GI:11605459
KEYWORDS Homo sapiens (human);
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 1363)
AUTHORS Liu,C., Zhang,L., Shao,Z.M., Beatty,P., Sartippour,M., Lane,T.F.,

```

```

Basky,S.H., Livingston,E. and Nguyen,M.
Identification of a novel endothelial derived gene EG-1
Biochem. Biophys. Res. Commun. 250 (1): 602-612 (2002)
2163837
PUBMED 1179215
REFERENCE 2 (bases 1 to 1661)
AUTHORS Nguyen,M.H. and Liu,C.
TITLE A Novel Gene as Tumor Angiogenesis Marker
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1661)
AUTHORS Nguyen,M.H. and Liu,C.
TITLE Direct Submission
JOURNAL Submitted 08 MAR 2001: Surgical Oncology, UCLA Medical School,
10835 Le Conte Ave., Los Angeles, CA 90095, USA
LOCATION/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
137..164
/contig_start:
/contig_end:
/contig_start_and_end_pos:marked
/protein_id="AAK12724.1"
/db_xref="GI:113605409"
/translations="MAAPLGCMFSGGPPGPPQAPGLPQASLLQAAAGCAPRPSSTL
VQLESSEFACFASLVSDYVNGTDOEERTVDDCCKFLDIARQTECFKCKLQLOL
SVQKSEVIEDVSELRNLRKQCALVCKELKLRHKQCVLEUINVQHKRPADIQGS
LAYEQSAN:PAPEKPT"
410 a 245 c 309 g 199 t
BASE COUNT 410 a 245 c 309 g 199 t
ORIGIN
Query Match 100.0%; Score 18; DB 9; Length 1363;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCACGTTGGCTTCAGAG 14
+ ||||| |||||
Db 549 TCACGTTGGCTTCAGAG 512

RESULT 10
AF321617/c 3109 bp mRNA linear PRI 07-SEP-2002
LOCUS Homo sapiens unknown mRNA.
DEFINITION AF321617
ACCESSION AF321617
VERSION AF321617.1 GI:11136411
KEYWORDS Homo sapiens (human);
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 3109)
AUTHORS Nguyen,M.H. and Liu,C.
TITLE Unpublished
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3109)
AUTHORS Nguyen,M.H. and Liu,C.
TITLE Direct Submission
JOURNAL Submitted 15 NOV-2000: Surgical Oncology, University of
California, Los Angeles, 10835 Le Conte Ave., Los Angeles, CA
90095, USA
LOCATION/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1759..2295
/codon_start:
/product="unknown"
/protein_id="AAG3862.1"
/db_xref="GI:11596416"
/translations="MAAPLGCMFSGGPPGPPQAPGLPQASLLQAAAGCAPRPSSTL
VQLESSEFACFASLVSDYVNGTDOEERTVDDCCKFLDIARQTECFKCKLQLOL
SVQKSEVIEDVSELRNLRKQCALVCKELKLRHKQCVLEUINVQHKRPADIQGS
LAYEQSAN:PAPEKPT"

```



Tue Oct 21 11:34:40 2003

CDS	46...585	/codon_start=1 /product="prion protein-like protein Dpl" /protein_id="AA032645.1" /db_xref="GI:16049298" /translation="MKNRLLGTWVAILCMLASHLSTVARGIKHFRFKWKKVLESSG GQITFARVAENRFGATIKQGRKLCIDFGAGNRYAANYWQFPIGYEGSEAVTK EMLVTSVNTATQANQAEKRSKQSKHQRVLRLIKIKCAKHCDFWLESCVALRV AVDQPMVCLLGFVWIVK"
polyA_site	17:4	
BASE COUNT	432 a 436 c 434 g 415 t	
ORIGIN	94.43; Score 17; EH 12; Length 1719; Best Local Similarity 100.0%; Pred. No. 88; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
CY	1	TCAGTTCGGCTCAG 17
DB	446	TCAATTTCATTCAG 330
RESULT 14		
BC025140	1929 bp mRNA linear RCD 20-SEP-2002	
LOCUS	Mus musculus, prion protein dublet, clone MGC334610 IMAGE160931, mRNA, complete cds.	
ACCESSION	BC025140	
VERSION	BC025140.1 GI:19263515	
KEYWORDS	MGC	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;	
AUTHORS	; (bases 1 to 1929)	
TITLE	Strausberg S.R.	
JOURNAL	Direct Submission	
REMARK	Submitted 05 MAR 2002; National Institutes of Health, Mammalian Gene Collection (MGC); Cancer Genomics Center, National Cancer Institute, 31 Center Drive, Room 1A43, Bethesda, MD 20892-2950, USA	
COMMENT	NIH-MGC Project URL: <a href="http://mgs.nci.nih.gov">http://mgs.nci.nih.gov</a> Contact: MGC help desk Email: <a href="mailto:exppcr@mail.nih.gov">exppcr@mail.nih.gov</a> Tissue Procurement: Gilbert Smith, Ph.D. cDNA Library Preparation: Life Technologies, Inc. DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL) Sequencing Center Center code: BCM-HGSC Web site: <a href="http://www.hgsc.bcm.tmc.edu/cdna/">http://www.hgsc.bcm.tmc.edu/cdna/</a> Contact: <a href="mailto:amg@bcm.tmc.edu">amg@bcm.tmc.edu</a> Gunaratne, P.H.; Garcia, A.M.; Lu, X.; Hulyk, S.W.; Hale, S.M.; Yoon, V.S.; Kowis, C.R.; Lawrence, S.; Martin, R.G.; Kozny, D.X.; Richards, S.; Gibbs, R.A.	
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/NCMI at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Series: IRAK Plate: 61 Row: d Column: 19 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA GI:16049298, location/Qualifiers ..1929 /organism="Mus musculus" /mol_type="mRNA" /strain="FVB/N" /db_xref="taxon:10090" /clone="MGC:35610 IMAGE:3600731" /issue_type="mammary tumor, Metathionien-TGF alpha model, 10 month old virgin mouse. Taken by biopsy." /clone_lib="NCI CGAP_Mam1" /lab_host="DH103" /note="Vector: pCMV-SPORT6"	
SOURCE		
CDS	47...581	/note="doppelganger; doppel" /codon_start=1 /product="prion protein-like protein Dpl" /protein_id="AA032544.1" /db_xref="GI:60492961" /translation="MKNRLLGTWVAILCMLASHLSTVARGIKHFRFKWKKVLESSG GQITFARVAENRFGATIKQGRKLCIDFGAGNRYAANYWQFPIGYEGSEAVTK EMLVTSVNTATQANQAEKRSKQSKHQRVLRLIKIKCAKHCDFWLESCVALRV AVDQPMVCLLGFVWIVK"
polyA_site	19:9	
CDS	44...581	/codon_start=1 /product="prion protein dublet" /protein_id="AA032542.1" /db_xref="GI:16049298" /translation="MKNRLLGTWVAILCMLASHLSTVARGIKHFRFKWKKVLESSG GQITFARVAENRFGATIKQGRKLCIDFGAGNRYAANYWQFPIGYEGSEAVTK EMLVTSVNTATQANQAEKRSKQSKHQRVLRLIKIKCAKHCDFWLESCVALRV AVDQPMVCLLGFVWIVK"
BASE COUNT	433 a 436 c 434 g 421 t	
ORIGIN	94.43; Score 17; EH 10; Length 1929; Best Local Similarity 100.0%; Pred. No. 88; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
CY	1	TCAGTTCGGCTCAG 17
DB	446	TCAATTTCATTCAG 330
RESULT 15		
BC025140	1929 bp mRNA linear RCD 16-OCT-1999	
LOCUS	Mus musculus, prion protein dublet-like protein Dpl, alternatively spliced mRNA, complete cds.	
ACCESSION	AF05165	
VERSION	AF05165.1 GI:16049295	
KEYWORDS	Mus musculus, house mouse	
SOURCE	Mus musculus	
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;	
REFERENCE	; (bases 1 to 1929)	
AUTHORS	Moore, R.C.; Lee, Y.; Silverman, G.L.; Harrison, P.M.; Strone, R.; Helinich, S.; Katsikatos, A.; Pasternak, S.H.; Chishti, M.A.; Liang, Y.; Vamvakopoulos, P.; White, K.; Smith, A.; Katsaris, S.; Carr-Saunders, A.; Torgov, E.; Finkler, S.B.; Wilson, D.W.; Trenchard, P.; Roodman, G.; and Westaway, D.	
TITLE	Axxin prion protein (PrP) solution time is associated with upregulation of the novel PrP-like protein Doppel	
JOURNAL	J. Mol. Biol. 281(4): 799-817 (1999)	
REMARK	2 bases 1 to 1929	
COMMENT	Direct Submission Submitted 10-JUL-1999; Institute for Neurodegenerative Dis-ease, HSE 7M1, 519 Parnassus Avenue, San Francisco, CA 94143, USA Location/Qualifiers ..1929 /organism="Mus musculus" /mol_type="mRNA" /db_xref="taxon:10090" /clone="MGC:35610 IMAGE:3600731" /issue_type="mammary tumor, Metathionien-TGF alpha model, 10 month old virgin mouse. Taken by biopsy." /clone_lib="NCI CGAP_Mam1" /lab_host="DH103" /note="Vector: pCMV-SPORT6"	

BASE COUNT 491 a 498 c 465 g 499 t  
GCTGCT

Query Match 94.4%; Score 17; DB 10; Length 19337  
Best Local Similarity 100.0%; Pred. No. 88;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CV 1 TCACGTTGCTTACAG 17  
D6 1 TCACGTTGCTTACAG 17  
442 TCACGTTGCTTACAG 17

Search completed: October 21, 2003, 21:41:46  
Job 1176 : 102.69 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

QM protein nucleic search, using frame\_plus\_p2n model

Run on: October 21, 2003, 09:21:19 , Search time 21 Seconds  
Without alignments:  
1106166 Million cell updates/sec

Titles: US-10-029-137-2  
Perfect score: 512  
Sequences: 1 XAFLGSGSEGPHTPPQAR.....GSLAYLQASANPAFLKPT 178

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 569978 segs, 220691566 residues  
Total number of hits satisfying chosen parameters: 1:19556

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0  
Maximum Match 1001  
Listing first 45 summaries

Command line parameters:  
-XDELFrame, p2n model, -DEV xlp  
-G/cgn2 1/USPTO specol\_p/US10029137/submit\_20102073\_111010\_6590/app.query.fasta\_1.327  
-DB Issued\_Patents NA -QFNT=lastap -SUFF=trim: MINMATCH=511 -LOOPCL=0  
-ICOPEXT 0 -UNITS=bits -STARTAL -END: 1 MATRX=Abacusmed2 -TPANS=human40.cdi  
-LIST:45 -DCALLION=200 -THR SCORE=pcr -THR MAX=100 -THR MIN=0 -ALIGN=15  
MODE=LOCAL -OUT=TMP=ptc -NCR=extl -HEAPSIZE=500 -MINDEX=0 -MAXLEN=200000000  
-USER=US10029137 -ACN 1 1 95 -ANALYT 20102001 17410 6690 -NCRU=6 -ICRU=3  
-NO WMAP -CARGUEERY -NEG SCORE=0 -WAIT -DSHLOCK=100 -LONGLOG  
-CPU TIMEOUT 120 -WARN TIMEOUT 10 -THRESHOLD 10 -XGAPOP 10 -XGAPEXT 0.5 -Fgapop=6  
-Ygapext=7 -Ygapop=10 -Ygapext 0.5 -Delop 6 -Delext 7

Database: Issued Patents NA  
1: /cgn2 6/prodata1/ina/5A COME seq  
2: /cgn2 6/prodata1/ina/5B COME seq  
3: /cgn2 6/prodata1/ina/6A COME seq  
4: /cgn2 6/prodata1/ina/6B COME seq  
5: /cgn2 6/prodata1/ina/6C COME seq  
6: /cgn2 6/prodata1/ina/6D COME seq

Find. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match	Description
1	98.5	10.7	1955 1 US-08-425-069-3 Sequence 3, Appl
2	98.5	10.7	1955 2 US-08-317-6445-3 Sequence 3, Appl
3	98.5	10.7	4295 4 US-09-620-3120-114 Sequence 38, App
4	98	10.7	3002 1 US-08-393-095-3 Sequence 3, Appl
5	96.5	10.5	1947 4 US-10-020-079-11 Sequence 11, Appl
6	96.5	10.5	1965 4 US-10-020-079-9 Sequence 9, Appl
7	96.5	10.5	2595 4 US-13-020-079-3 Sequence 3, Appl
8	96.5	10.5	2613 4 US-10-020-079-1 Sequence 10, Appl
9	96.5	10.5	2870 4 US-10-020-079-19 Sequence 10, Appl
10	96.5	10.5	2888 4 US-10-020-079-17 Sequence 17, Appl
11	96.5	10.5	2856 4 US-10-020-079-35 Sequence 35, Appl
12	96.5	10.5	2874 4 US-10-020-079-33 Sequence 33, Appl

13	96.5	10.5	2931 4 US-10-020-079-27 Sequence 27, Appl
14	96.5	10.5	2949 4 US-10-020-079-25 Sequence 25, Appl
15	95.5	10.4	1840 4 US-09-016-434-1367 Sequence 1367, Ap
16	95.5	10.4	3174 4 US-09-489-847-86 Sequence 86, Appl
17	95	10.3	5115 3 US-09-348-518C-3 Sequence 3, Appl
18	95	10.3	5115 3 US-09-476-509B-3 Sequence 3, Appl
19	91	9.9	876 4 US-03-252-991A-8476 Sequence 8476, Ap
20	91	9.9	1776 4 US-09-252-991A-8333 Sequence 8333, Ap
21	89.5	9.7	4031 1 US-08-159-784-1 Sequence 1, Appl
22	88.5	9.6	3402 4 US-09-252-991A-15560 Sequence 15560, A
23	88.5	9.6	3606 4 US-09-252-991A-15689 Sequence 15689, A
24	88	9.6	1098 4 US-09-252-991A-15756 Sequence 15756, A
25	87	9.5	1605 4 US-09-252-991A-13294 Sequence 13294, A
26	87	9.5	1941 4 US-03-252-991A-12669 Sequence 12669, A
27	87	9.5	2967 4 US-09-252-991A-13128 Sequence 13128, A
28	86.5	9.4	1470 4 US-09-252-991A-369 Sequence 369, App
29	86.5	9.4	7419 4 US-09-252-991A-481 Sequence 481, App
30	86.5	9.4	7449 4 US-09-252-991A-396 Sequence 396, App
31	86	9.4	549 3 US-08-851-19C-2 Sequence 2, Appl
32	86	9.4	3711 4 US-09-883-134-3 Sequence 3, Appl
33	86	9.4	4034 4 US-03-883-134-5 Sequence 5, Appl
34	85.5	9.3	2343 4 US-09-643-597-368 Sequence 368, App
35	85.5	9.3	4849 4 US-09-620-312D-39 Sequence 39, Appl
36	85	9.3	495 4 US-09-252-991A-3717 Sequence 3717, Ap
37	85	9.3	2766 4 US-09-252-991A-3756 Sequence 3756, Ap
38	85	9.3	2874 4 US-09-252-991A-3837 Sequence 3837, Ap
39	85	9.3	4401765 3 US-09-103-840A-2 Sequence 2, Appl
40	85	9.3	441529 3 US-09-103-840A-1 Sequence 1, Appl
41	84.5	9.2	2680 2 US-08-533-306A-5 Sequence 5, Appl
42	84.5	9.2	2680 2 US-08-742-923A-5 Sequence 5, Appl
43	84.5	9.2	2887 2 US-08-533-306A-3 Sequence 3, Appl
44	84.5	9.2	2887 2 US-08-742-923A-3 Sequence 3, Appl
45	84.5	9.2	35081 2 US-08-752-76CA-1 Sequence 1, Appl

ALIGNMENTS

RESULT 1  
US-08-425-069-3  
Sequence 3, Application US/08425069  
Patent No. 572881C  
GENERAL INFORMATION:  
APPLICANT: Lewis, Randolph V.  
APPLICANT: Xu, Ming  
TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK  
TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL  
TITLE OF INVENTION: CONTAINING THE ISOLATED ENA, AND PRODUCTS THEREOF  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolach & Birch  
STREET: 301 No. 5728810th Washington Street  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22046  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/425,069  
FILING DATE: 19-APR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald W  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 447-106P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 205-8000  
TELEFAX: (703) 205-8050  
TELEX:

INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1995 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 HYPOTHETICAL: NO  
 IMMEDIATE SOURCE:  
 CLONE: p6B  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..1785  
 US-08-425-069-3

Alignment Scores:  
 Pred. No.: 0.474 Length: 1995  
 Score: 98.50  
 Percent Similarity: 40.21%  
 Best Local Similarity: 26.80%  
 Query Match: 10.73%  
 Gaps: 9

US-10-029-137-2 (1-178) x US-08-425-069-3 (1-1995)

QY 6 GlyGlyMetPheSerGlyGlnProPro-----GlyProGlnAlaProPro 21  
 DB 415 GGAGTTATGACCAAGGTCACCAAGGTCCTGGAGGATATGACCAAGGACACAGGACCA 474  
 QY 22 GlyLeuProGlyGlnAlaSerLeuLeuLeuGlnAlaAla----- 33  
 DB 475 TCTGACACAGGTAGTGGCTGGAGGAGGCGCTGTCGATCAGGACCTGGACACACAGCA 534  
 QY 34 -----ProGlyAlaProArg 38  
 DB 535 CCAGACGACATGACCAAGGTCACCAAGGTCCTGGAGGATATGACCAAGGACACAGCA 594  
 QY 39 ProSerSerSerThrLeuValaspGluLeuGluSerPheGluAlaCysPheAlaSer 58  
 DB 535 CCATCTGACCAAGGT-----AGTGGCGCTGACGCGCC 627  
 QY 59 LeuValSerGlnAspTyrValasnGlyThrAspGlnGluGluIleArgThrGlyValas 78  
 DB 628 GCGGACGACATCAGGAC-----CTGGACACACAGGACACAGGACCA 672  
 QY 78 PGINcysTleGlnLysPheLeuAspTleAlaArgGlnThrGluCysPhePheLeuGlnLy 98  
 DB 673 CCAG --GTCACACAGGTCGAGGATTATGGACACAGCAACACAGGACTATCTGGACCA 729  
 QY 98 sArgLeuGlnLeuSerValGlnLysProGluGlnValle-----LysGluAs 114  
 DB 730 GGCACGTCAGCTG-----CAGCAGCGGACGACGACGACGACGACGACGACGACG 783  
 QY 114 pValSerGluLeuArgAsnGluLeuGlnArgLysAspAlaLeuValGlnLysHisLeuTh 134  
 DB 784 TATGACACAG -----GACACACAGGACCTCTGAGCGCGGTAGTGGCGCT 825  
 QY 134 LysLeuAlaGlnHisTyrpGlnGlnValleGluAspTleAsnValGlnHisLysLysProAl 154  
 DB 829 GCACGACGACGCGCGCGGACGACGACGACGACGACGACGACGACGACGACGACGACG 885  
 QY 154 aAspTlePheGlnGlySerLeuAlaTyrLeuGluGlnAla 167  
 DB 889 GGATATGACGACGACGACACACAGGACCATCTGGACGACGCA 928

# RESULT 2

US-08-317-844B-3  
 Sequence 3, Application US/08317844B  
 Patent No. 5989894  
 GENERAL INFORMATION:  
 APPLICANT: Lewis, Randolph V.  
 APPLICANT: Xu, Ming  
 APPLICANT: Hinman, Michael B.

TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK  
 TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL  
 TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF  
 NUMBER OF SEQUENCES: 67  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Smith, Stewart, Kolesch & Smith  
 STREET: 301 No. 598-44th Washington Street  
 CITY: Falls Church  
 STATE: Virginia  
 COUNTRY: U.S.A.  
 ZIP: 22046  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC DOS/MS-DOS  
 SOFTWARE: Patent Invention #100, Division #1125  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: 08/117,640  
 FILING DATE: 4/10/94  
 CLASSIFICATION: 415  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Mulvey, James, G. & Co.  
 REGISTRATION NUMBER: 8,977  
 REFERENCE/CIKNET NUMBER: 1447 125P  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 703/241-1300  
 TELEFAX: 703/241-2818  
 TELEX: 245145

INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1995 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 HYPOTHETICAL: NO  
 IMMEDIATE SOURCE:  
 CLONE: p6B  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..1785  
 US-08-417-844B-3

Alignment Scores:  
 Pred. No.: 0.474 Length: 1995  
 Score: 98.50  
 Percent Similarity: 40.21%  
 Best Local Similarity: 26.80%  
 Query Match: 10.73%  
 Gaps: 9

US-10-029-137-2 (1-178) x US-08-317-844B-3 (1-1995)

QY 6 GlyGlyMetPheSerGlyGlnProPro-----GlyProGlnAlaProPro 21  
 DB 415 GGAGTTATGACCAAGGTCACCAAGGTCCTGGAGGATATGACCAAGGACACAGGACCA 474  
 QY 22 GlyLeuProGlyGlnAlaSerLeuLeuLeuGlnAlaAla----- 33  
 DB 475 TCTGACACAGGTAGTGGCTGGAGGAGGCGCTGTCGATCAGGACCTGGACACACAGCA 534  
 QY 34 -----ProGlyAlaProArg 38  
 DB 535 CCAGACGACATGACCAAGGTCACCAAGGTCCTGGAGGATATGACCAAGGACACAGCA 594  
 QY 39 ProSerSerSerThrLeuValaspGluLeuGluSerPheGluAlaCysPheAlaSer 58  
 DB 535 CCATCTGACCAAGGT-----AGTGGCGCTGACGCGCC 627  
 QY 59 LeuValSerGlnAspTyrValasnGlyThrAspGlnGluGluIleArgThrGlyValas 78  
 DB 628 GCGGACGACATCAGGAC-----CTGGACACACAGGACACAGGACCA 672





```

1 APPLICATION NUMBER: US/08/393,985
2 FILING DATE:
3 CLASSIFICATION: 435
4 ATTORNEY/AGENT INFORMATION:
5 NAME: Sholtz, Charles K.
6 REGISTRATION NUMBER: 38,615
7 REFERENCE/DOCKET NUMBER: 8600-0152
8 TELECOMMUNICATION INFORMATION:
9 TELEPHONE: 315-324-6880
10 TELEFAX: 315-324-0960
11 INFORMATION FOR SEQ ID NO: 1:
12 SEQUENCE CHARACTERISTICS:
13 LENGTH: 3000 base pairs
14 TYPE: nucleic acid
15 STRANDEDNESS: double
16 TOPOLOGY: unknown
17 MOLECULE TYPE: DNA TO RNA
18 HYPOTHETICAL: N
19 ANTI-SENSE: N
20 ORIGINAL SOURCE:
21 INDIVIDUAL ISOLATE: Rat lymphatic B-lymphoma Mammals
22 FEATURE:
23 NAME/KEY: CDS
24 LOCATION: 213..1077
25 US-08 393 985-3

```

```

Alignment Scores:
Pred. No.: 1000
Score: 98.00
Percent Similarity: 43.45%
Best Local Similarity: 27
Mismatch: 61
Query Match: 10,681
DB: 1

```

```

US-10 029-137 2 (1-178) x US-08 393-985 3 (1-3000)
QY 13 ProProGlyProGlyLeuAlaProGlyLeuProGlyGlnAlaSerLeuLeuGlnAla 17
DB 65 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 169
QY 33 AlaProGlyAlaProArgProSerSerSerThrLeuValAspLeuLeuSerLeuPhe 37
DB 110 TACCGGCTTCTCCAGACCTCCGCGCTCCAGACGAGGCTAGACGAGGCTAGACGAGG 163
QY 53 GlnAlaCysPheAlaSerLeuValSerGlnAspTyrValAspGlyThrAsp 69
DB 192 GAG 192
QY 70 GlnGluGluGluLeu-----ArgThrGlyValAspGlnCys-----LeuGlnLysPhe 84
DB 191 GAGGAGGAGGCTTCGCGAGGATGAAGGATCGGAGCTCAGAGCTCGGAGCTGAAGAGCA 250
QY 85 LeuAspLeuAlaArgGlnThrGluCysPhePheLeuGlnLysArgLeuGlnLeuSerVal 104
DB 251 GTGACGATGAAGAGGAG-----TGTTTACTTGATCAGAGCGACTTATGAGAG 301
QY 105 GlnLysProGlnValLys-----GlnAspValPheThrLeuAla 119
DB 302 AGTTCTTTGAGCAGGCTGGAAGAGATCCGAGCTTCATCGCAAACTGTCGAGGATG 360
QY 120 AsnGlnLeuGlnAspLysAspAlaLeuValGlnLysHisLeuThrLysLeuArgHisTyr 139
DB 361 GAGCAAGTGAAGAACAGCAGCAGCTCCATTTCTGTTCTCCGCAACCCGATGAGAGACT 420
QY 140 GlnGlnValLeuGlnLysPheAsnValGlnHisLysPheGlnAlaSerProGly 154
DB 421 AAGCAGGAGCTGGAGGAGCTTCAAGGAGAGATCAAAAGAGAGAGAGAGAGAGAGAGAG 480
QY 160 SerLeuAlaTyrLeuGlnGlnAla 167
DB 481 AAGTTGAAGGATCCGAGGAGAG 504

```

RESULT 5

```

US-10 029-137 1
1 Sequence ID: Application US/08/393,985
2 Patent No: 6579710
3 GENERAL INFORMATION:
4 APPLICANT: Turner, C. Alexander Jr.
5 APPLICANT: Mathai, Brian
6 APPLICANT: Prasad, Carl John
7 TITLE OF INVENTION: NO. 6579710, NO. 6579711, NO. 6579712, NO. 6579713
8 FILE REFERENCE: US/08/393,985
9 CURRENT APPLICATION NUMBER: US/08/393,985
10 CURRENT FILING DATE: 2003-12-10
11 PRIOR APPLICATION NUMBER: US/08/393,985
12 PRIOR FILING DATE: 2000-12-10
13 PRIOR APPLICATION NUMBER: US/08/393,985
14 PRIOR FILING DATE: 2001-01-04
15 NUMBER OF SEQ ID NOS: 43
16 SOFTWARE: PARSED FOR CDS, GENE, AND ORF
17 SEQ ID NO: 1
18 TYPE: CDS
19 ORGANISM: Rat
20 US-10 029 137-2

```

```

Alignment Scores:
Pred. No.: 1000
Score: 98.00
Percent Similarity: 43.45%
Best Local Similarity: 27
Mismatch: 61
Query Match: 10,681
DB: 1

```

```

US-10 029-137 2 (1-178) x US-08 393-985 3 (1-3000)
QY 2 AlaAlaProGlyProGlyLeuAlaProGlyLeuProGlyGlnAlaSerLeuLeuGlnAla 17
DB 109 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 162
QY 14 ProProGlyProGlyLeuAlaProGlyLeuProGlyGlnAlaSerLeuLeuGlnAla 17
DB 163 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 169
QY 33 AlaProGlyAlaProArgProSerSerSerThrLeuValAspLeuLeuSerLeuPhe 37
DB 223 GAGGAGGAGGCTTCGCGAGGATGAAGGATCGGAGCTCAGAGCTCGGAGCTGAAGAGCA 250
QY 38 AlaProSer-----ArgThrGlyValAspGlnCys-----LeuGlnLysPhe 84
DB 283 CTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 442
QY 51 SerPheGlnAlaCysPheAlaSerLeuValSerGlnAspTyrVal-----LeuGlnLys 65
DB 349 ATCGCGAGGAGGCTTCGCGAGGATGAAGGATCGGAGCTCAGAGCTCGGAGCTGAAGAGCA 402
QY 65 LeuAspLeuAlaArgGlnThrGluCysPhePheLeuGlnLysArgLeuGlnLeuSerVal 104
DB 403 AAGTTTGAAGGATCCGAGGAGAGATCCGAGCTTCATCGCAAACTGTCGAGGATG 360
QY 85 GlnLysProGlnValLys-----GlnAspValPheThrLeuAla 119
DB 402 AGTTCTTTGAGCAGGCTGGAAGAGATCCGAGCTTCATCGCAAACTGTCGAGGATG 360
QY 120 AsnGlnLeuGlnAspLysAspAlaLeuValGlnLysHisLeuThrLysLeuArgHisTyr 139
DB 547 AAGTTGAAGGATCCGAGGAGAG 504
QY 127 SerLeuAlaTyrLeuGlnGlnAla 167
DB 604 TTTAAGTATAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 663

```





US 10,020,039-19

Alignment Scores:

Pred. NC.

Score:

Percent  
Sulfur

345. local. similar.

Curry, Mad. Co.

De:  
Cody + Co, Inc.

Length: 2:570

Matches: 47

Conservative: 25

M: s-a-chos:

6:51:30

[illegible]

US-10	020	27.2	11.78	$\times 10^{-5}$	0.20	0.79	1.0	0.69
-------	-----	------	-------	------------------	------	------	-----	------

Cy	2	AlaAlaProLeuGlyIleValPheSerHisLysPro.....	139
D6	109	GCGGGCCGCCGGAGAG...AGTGGACCAACCAGCAGCATTCAGCCGCCGCCGC	162
Cy	14	-----ProGlyProThrGlnAfaProThrGlyLeuProGlyIysNAlasrLeu.....	23
D6	143	CCGCGCCCCTCCGGGAGATCGAGCGCGCGAGCTGTGTCTCGCCATCGCGCGCACGC	222
Cy	30	-----LeuGlnAalaaPheGlyValaPro37	
D6	223	GGACCGCGAGCGGACACAGAGGTATATGGCCCCCTCAGCGCAGSCCGCGGGACACG	282
Cy	38	ArgProser-----SerSerThrLeuvalAspGluLeuGluSer50	
D6	293	CCTCCCTCTGGTGCGUGAGAGATGATTATCGCGCGCCCTAAAGSAGGAACAACATG	342
Cy	51	SerPheGluAlacysPheAlaserLeuValSerGlnAspTyVal.....	65
D6	343	ASTGGGGAGGAGGAGTAGCGCGACACTATATCGCGTCAACTACGTGTCAAGGATCGCTG	402
Cy	66	-----AsnAlaThrAspGlnGluIleValThrGlyThrgLyvalAspGln79	
D6	403	AAGGTGCTGAAAAATAATGAGAGCGGCTTTCTGAGACTACAGGTCATGGACCTG	462
Cy	80	CysIleGlnLysPheLeuAspLysAlaArgGlnThrGluCysPhePheLeuGlnLysArg93	
D6	463	CTGACCCAGGGAG....AATTTGGCCCTCAAGGTGGAG.....	495
Cy	100	LeuGlnLeuSerValGlnLysProGlnIleValIleLysGluAspValSerGluLeuArg129	
D6	496	-----TCAGCCAGGAGCTCAGGAGGTCCTTAAGAATGAGGTGGCGTGGCTCAAG	546
Cy	120	AsnGlnLeuGlnArgLysAsp.....	126
D6	547	---AAGTTGCAAGGAAATGATATGTTGTGCAATTCATGTTGTGGCAGAACGAGAG	603
Cy	129	-----AlaLeuValThrPheAlaMetHisLysPheArgHisThrTrpGln140	
D6	604	TTTAACTATGTATATAGATATTTTATATATATATATATATATATATATATATAT	653
Cy	141	-----GlnValLeuGluLasp145	
D6	664	CCGCGAGGCACCTTCACTCTGATACACCAATATGCGSTCGTAAGCAGATCTGGAGTCC	723
Cy	146	IleAsnValGlnHisLys.....	157
D6	724	ATCAGAGCCATCCACTCTGTGGCTCTCTCTATACCGGTGACATTAAGCTTCAAACCTTGGC	783
Cy	158	GlnGlySerLeu161	
D6	784	ATGGCAGGCTG795	

## RESULTS

US-10-020-329-17

Sequence 17. Application JS/:00200379

Parcel No. 65797:0

GENERAL INFORMATION:

APPLICANT: Turner, C. Alexander, Jr.

APPLICANT: Marbury, Brian

APPLICANT: Macaulay, Brian  
APPLICANT: Erididdle, Carl  
APPLICANT: Erididdle, Carl

APPLICANT: FIDUE, Carl Johan  
TITLE OF INVENTION: No. 657971C: Human Kinases and polynucleotides encoding the same



```

Db 411 AAGTGTCTGAAAAGATCGGGGCGGGGCTTTTGGTGGAGAGCTAGAGAGCATGACCTG 480
QY 60 CysLeuGlnLysPheLeuAspIleAlaArgGlnThrGlnCysPhePheLeuGlnLysArg 99
Db 481 CTACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 513
QY 100 LeuGlnLeuSerValGlnLysPheGlnLysValIleLeuGlnLysValSerGlnLeuArg 119
Db 514 TCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 564
QY 120 AsnGlnLeuGlnLysArgGlnLysArgGlnLysArgGlnLysArgGlnLysArgGlnLys 128
Db 565 AAGTGTCTGAAAAGATCGGGGCGGGGCTTTTGGTGGAGAGCTAGAGAGCATGACCTG 621
QY 127 TTTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 642
Db 642 TTTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 681
QY 141 TTTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 145
Db 682 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 741
QY 146 IleAsnValGlnLysArgGlnLysArgGlnLysArgGlnLysArgGlnLysArgGlnLys 157
Db 742 ATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 801
QY 159 GlnGlySerLeu 161
Db 802 ATGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 811

```

## RESULT 1:

US-10-029-079-25

Sequence 25, Application US/100/000099

Parent No. 6579710

GENERAL INFORMATION:

APPLICANT: Turner, C. Alexander et al.

APPLICANT: Nachar, Brian

APPLICANT: Fiedler, Carl Johan

TITLE OF INVENTION: No. 6579710: Human Kinases and Polymorphic Residues Encoding the Same

FILE REFERENCE: LEX-0281-USA

CURRENT FILING DATE: 2001-12-12

PRIOR APPLICATION NUMBER: US/100/029,079

PRIOR FILING DATE: 2000-12-12

PRIOR APPLICATION NUMBER: US 60/289,422

PRIOR FILING DATE: 2001-05-08

NUMBER OF SEQ ID NOS: 40

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 25

TYPE: DNA

ORGANISM: Homo sapiens

US-10-029-079-25

## Alignment Scores:

Pred. No.: 138

Score: 96.50

Percent Similarity: 33.61%

Best Local Similarity: 23.36%

Query Matches: 10.51%

DB: 4

US-10-029-137-2 (1-178) x US-10-029-079-25 (1-2941)

```

QY 2 AlaAlaProLeuGlyGlyMetPheSerGlySerPro 13
Db 109 GCGGCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 162
QY 14 -----ProGlyProProGlnAlaProPheSerGlySerPro 29
Db 163 GCGGCGCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 222
QY 30 -----LeuGlnAlaLeuPheGlyAlaPro 37

```

```

Db 223 GCGGCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 282
QY 38 ArgProSer 50
Db 283 CCGGCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 342
QY 51 SerProGlnAlaLysPheLeuValSerGlnLysArgGlnLysArgGlnLysArgGlnLys 65
Db 143 AGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 402
QY 56 AsnGlyThrAspGlnGlnLysArgGlnLysArgGlnLysArgGlnLysArgGlnLysArg 79
Db 403 AAGGCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 462
QY 80 CysLeuGlnLysPheLeuAspIleAlaArgGlnThrGlnCysPhePheLeuGlnLysArg 99
Db 463 CTACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 495
QY 100 LeuGlnLeuSerValGlnLysPheGlnLysValIleLeuGlnLysValSerGlnLeuArg 119
Db 496 TCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 546
QY 120 AsnGlnLeuGlnLysArgGlnLysArgGlnLysArgGlnLysArgGlnLysArgGlnLys 128
Db 547 AAGTGTCTGAAAAGATCGGGGCGGGGCTTTTGGTGGAGAGCTAGAGAGCATGACCTG 603
QY 127 AlaLeuValGlnLysHisLeuThrLysLeuArgHisThrProGln 140
Db 604 TTTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 663
QY 141 IleAsnValGlnLysArgGlnLysArgGlnLysArgGlnLysArgGlnLysArgGlnLys 145
Db 664 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 723
QY 146 IleAsnValGlnLysArgGlnLysArgGlnLysArgGlnLysArgGlnLysArgGlnLys 157
Db 724 ATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 783
QY 159 GlnGlySerLeu 161
Db 784 ATGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 795

```

## RESULT 14:

US-10-029-079-25

Sequence 25, Application US/100/000099

Parent No. 6579710

GENERAL INFORMATION:

APPLICANT: Turner, C. Alexander et al.

APPLICANT: Nachar, Brian

APPLICANT: Fiedler, Carl Johan

TITLE OF INVENTION: No. 6579710: Human Kinases and Polymorphic Residues Encoding the Same

FILE REFERENCE: LEX-0281-USA

CURRENT FILING DATE: 2001-12-12

PRIOR APPLICATION NUMBER: US/100/029,079

PRIOR FILING DATE: 2000-12-12

PRIOR APPLICATION NUMBER: US 60/289,422

PRIOR FILING DATE: 2001-05-08

NUMBER OF SEQ ID NOS: 40

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 25

TYPE: DNA

ORGANISM: Homo sapiens

US-10-029-079-25

## Alignment Scores:

Pred. No.: 139

Score: 96.50

Percent Similarity: 33.61%

Best Local Similarity: 23.36%

Query Matches: 10.51%

Length: 2949

Matches: 57

Conservative: 25

Mismatches: 63

Indels: 59





```

QY      143 LeuGluAspIleAsnValGlnHisLysLys----- --ProAlaAspIlePro- - 157
DB      143 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      1502 CTCAGGACATC---ATCGAGACACTCAATACGTCGGGGAGCCGCCCGACAGACAGTGCAC 1558
QY      158 ----- ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      158 ----- ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      1559 CCACTGAGGACATCTCGAAATCTCCATATTCACACTGCAATCATGTGCACTGGATCGGAC 1618
QY      166 GlnAlaSerAla 169
DB      166 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      1619 CAGAACTCGGCC 1639

```

Search completed: October 21, 2003, 01:40:14  
 Job time: 03 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Computer Ltd.

CM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 21, 2003, 00:17:04 : Search time 2015 Seconds  
(without alignments)  
2125.895 Million cell updates/sec

Title: US-10-029-137-2  
Perfect score: 918  
Sequence: 1 MAPUGMSSGSRGPRQDA.....GLAY:EWASNPAT:KPT 178

Scoring table:  
RGSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delcxt 7.0

Searched: 2278192 seqs, 1252238056 residues  
Total number of hits satisfying chosen parameters: 45562764

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame+p2nmodel -DEV=xip  
-Q/fgene-1/USPto-spool-p/US10029137/total-01102003-171430-6676/app-query-fasta-1.1327  
-DB=EST -QFMT=fastap -SUFFIX=ext -V=NMATCH=21 -LQOP=2.0 -LQOPEXT=0  
-UNITS=hits -START=1 -END=1 -MATRIX=blockm2 -TRANS=human42 -cdi -LIST=45  
-LOCALIGN=200 -THR SCORE=pct -THR YAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NCRM=ext -HEARSIZE=500 -MINLEN=0 -MAXLEN 2000000000  
-USER US10029137 -ACGN 1.1 -8874 -strand 20102003-171430-6676 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSEBLOCK=100 -LOCKLOG  
-DEV TIMECUT=120 -WARN TIMECUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6  
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -DELCP=6 -DELEXT=7

Database : EST:  
1: em\_estba:  
2: em\_estbm:  
3: em\_estbi:  
4: em\_estbu:  
5: em\_estov:  
6: em\_estpl:  
7: em\_estli:  
8: em\_intli:  
9: gb\_est1:  
10: gb\_est2:  
11: gb\_hic:  
12: gb\_est3:  
13: gb\_est4:  
14: gb\_est5:  
15: em\_estiua:  
16: em\_estom:  
17: em\_gss\_fur:  
18: em\_gss\_inv:  
19: em\_gss\_pln:  
20: em\_gss\_vit:  
21: em\_gss\_fun:  
22: em\_gss\_mus:  
23: em\_gss\_rus:  
24: em\_gss\_pro:  
25: em\_gss\_red:  
26: em\_gss\_phg:  
27: em\_gss\_vrl:  
28: gb\_gss1:

29: gb\_gss2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	918	100.0	693	10	BG473678
2	916	100.0	728	10	BF977554
3	918	100.0	788	10	BF440088
4	918	100.0	807	14	CB992984
5	918	100.0	831	10	BG141322
6	918	100.0	879	13	BU158756
7	918	100.0	886	14	CD389145
8	918	100.0	1019	13	EX437420
9	918	100.0	1077	11	AF118059
10	918	100.0	1173	13	EX376388
11	915	99.7	881	13	BU176649
12	913	99.5	783	14	CD346323
13	913	99.5	873	13	CG961556
14	913	99.5	1074	10	BF381359
15	912	99.3	1201	13	BM385486
16	911	99.2	630	12	BM789663
17	909	99.0	739	10	BG480455
18	909	99.0	1055	13	BX345448
19	907	98.8	911	13	BQ425134
20	906	98.7	1201	13	BX447239
21	905	98.6	567	14	CB160819
22	899	97.9	573	14	CA952729
23	898	97.8	755	13	BU183150
24	898	97.8	792	10	BG685333
25	897	97.7	559	9	A1929383
26	896	97.6	744	10	BG733190
27	893	97.3	573	14	CB160523
28	892	97.2	925	10	BF381358
29	888	96.7	810	12	BF171096
30	883	96.2	953	10	BF205933
31	879	95.8	522	14	CB178337
32	877	95.5	734	12	BU586759
33	877	95.3	815	13	BU855470
34	877	95.3	827	13	BJ521337
35	877	95.3	1442	10	BG141321
36	876	95.4	1020	13	BX437419
37	871	94.9	509	12	BM72447
38	871	94.9	1178	13	BX357725
39	870	94.8	591	13	BU308395
40	870	94.8	650	14	BY725533
41	870	94.8	666	12	B1648127
42	870	94.8	671	14	BY747831
43	870	94.8	696	14	CB596854
44	870	94.8	739	14	CA465315
45	870	94.8	748	13	BU847705

ALIGNMENTS

RESULT :  
BG473678  
LOCUS : 602515845F1 NIH\_MGC\_16 Homo sapiens cDNA clone IMAGE:4547663 5' .  
DEFINITION : mRNA sequence.  
ACCESSION : BG473678  
VERSION : BG473678.1 GI:13425955  
KEYWORDS : EST.  
SOURCE : Homo sapiens (human).  
ORGANISM : Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE : (Bases 1 to 693).



```

162 ACCAGTACTTGGTGGACAGTGGGATGATCTTCGAGGCTTCCTTCATCTCTGGTG 221
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
61 SerGlnAspTyrValAsnGlyThrAspGlnGluGluIleAsnThrGlyValAspGlnCys 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
272 AGTCAGGACTATGTAATGACACCGATACGAGAAATTCGACCGGTGTGATCAGTGT 281
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
a1 5'cGlnGlySerPheLeuArgIleAlaAlaGlnThrGluCysPhePheLeuGlnGlyArgGlu 100
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
292 ACCGAGAGCTTCTGGATATCAAGAGAAGAGAAAGTCTTCTTACAAAAGAGTTG 341
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
101 GlnGluSerValGlnGlySerProGluGluValSerGlnValSerGlnGluArgGln 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
142 CAGTATATGTCAGAGAAATCAATCAATATCAAGAGAGAGATATTCAGAACTAAGCAAT 401
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
111 GlnGluGlnArgGlyAsnAlaIleValAlaIleValIleValIleValIleValIleVal 140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
402 GAATACACGGGAGAGATGCAATACATGAGAGAGAGTTCAGAGAGAGAGATGCGAG 461
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
141 GlnValLeuGlnAspIleAsnValAlaGlnGlnGlySerPheAlaAspIlePheGlnGlySer 160
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
462 CAGGTCTCTGGAGGAGATCAACCTGAGAGACACAAAGAGCCGCGCATCTCCAGAGTCC 521
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
161 LeuAlaTyrLeuGlnGlnAlaSerAlaAsnIlePheAlaArgGlnGlySerProThr 178
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
522 TGGGCTACTGATGAGAGATATGAGAGATATGAGAGATATGAGAGATATGAGAGATATGAGAG 575
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3
LOCUS      BF440088      587 bp     mRNA     linear     EST 01-NOV-2003
DEFINITION M414 HVEC cDNA library Homo sapiens cDNA, mRNA sequence.
ACCESSION  BF440088
VERSION     BF440088.1 GI:11467963
KEYWORDS   EST.
SOURCE      Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 bases 1 to 587.
REFERENCE  1 (bases 1 to 587)
AUTHORS   Nguyen M.H. and Liu C.H.
TITLE     Identification of Genes in HVECs
JOURNAL   Unpublished
COMMENT   Contact: Mai H. Nguyen, Gishui Liu
            University of California Los Angeles
            4417 CHS, Surgery Oncology, UCLA, 15813 Wilshire Ave., Los Angeles,
            CA 90095, USA
            Tel: 310 267 1961
            Fax: 310 825 7579
            Email: maihu@ucla.edu, liuc@ucla.edu
PCR PRIMER  FORWARD: 77
            REVERSE: 77
            Insert Length: 788      Std Error: 0.05
            Seq primer: T7 and T3
POLY(A)=Yes
FEATURES             Location/Qualifiers
     source            1..788
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /cell_line="Human Umbilical Vein Endothelial Cell"
                        /clone_lib="HVEC cDNA library"
                        /note="Vector: Lambda UNIZAP Vector, Site 1: EcoRI;
                        Site 2: XhoI"
     BASE COUNT       216 a 179 c 144 g 199 t
     ORIGIN
Alignment Scores:
Pred. No.:      8,224-71
Score:          916.00
Percent Similarity: 100.00%
Conservative: 100.00%
Best Local Similarity: 100.00%
Mis-matches: 0

```

```

Query Match:      100.00%      Indels:
DB:                10          Gaps:      0
US-10-029-137-2 (1:178) x BF440088 (1:788)
QY      1 MetAlaAlaProGlnGlyGlyMetPheSerGlyGlnProProGlyProGlnAlaPro 20
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      12 ATGGCGGTGTCCTACTAGGGGGTATGTTTCTGGCGGAGCCACCGGTCTCCCTCCAGGCCCG 71
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      21 ProGlyLeuProGlyGlnAlaSerLeuLeuGlnAlaAlaProGlyAlaProAspProSer 40
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      72 CGGGGATTTCCGGGCGCAAGTTCGTTCTTTCAGGCAGCTCCAGCGCTCTCCAGACCTTCC 131
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      41 SerSerThrLeuValAspGlnLeuGlnSerSerPheGlnAlaCysPheAlaSerLeuVal 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      132 ASCAGTATATTTGGTGACAGAGTGGAGTCATCTTTCGAGGCTTCTTCATCTCTCTGATG 191
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      61 SerGlnAspTyrValAsnGlyThrAspGlnGluGluIleAsnThrGlyValAspGlnCys 80
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      192 AGTCAGGACTATGTAATGACACCGATACGAGAAATTCGACCGGTGTGATCAGTGT 251
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      81 5'cGlnGlySerPheLeuArgIleAlaAlaGlnThrGluCysPhePheLeuGlnGlyArgGlu 100
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      252 AGCCAGATTTCTGGATATCAAGAGAAGAGTTCAGAGAGAGATATTCACAAAAGAGTTG 311
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      101 GlnGluSerValGlnGlySerProGlnValIleValSerGlnValSerGlnGluArgGln 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      312 CAGTATATGTCAGAGAAATCAACCTGAGAGACACAAAGAGCCGCGCATCTCCAGAGTCC 371
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      121 GlnValLeuGlnAspIleAsnValAlaGlnGlnGlySerPheAlaAspIlePheGlnGlySer 140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      372 GAATACACGGGAGAGATGCAATACATGAGAGAGAGTTCAGAGAGAGAGATGCGAG 431
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      141 GlnValLeuGlnAspIleAsnValAlaGlnGlnGlySerPheAlaAspIlePheGlnGlySer 160
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      432 CAGGTCTCTGGAGGAGATCAACCTGAGAGACACAAAGAGCCGCGCATCTCCAGAGTCC 491
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      161 LeuAlaTyrLeuGlnGlnAlaSerAlaAsnIlePheAlaArgGlnGlySerProThr 178
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      492 TGGGCTACTGATGAGAGATATGAGAGATATGAGAGATATGAGAGATATGAGAGATATGAGAG 545
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4
LOCUS      CB992984      507 bp     -RNA     linear     EST 01-MAY-2003
DEFINITION AGENCOURT 135189822 NIH MGC 148 Homo sapiens cDNA clone
ACCESSION  CB992984
VERSION     CB992984.1 GI:30297504
KEYWORDS   EST.
SOURCE      Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 bases 1 to 507.
REFERENCE  1 (bases 1 to 507)
AUTHORS   NIH MGC http://mgi.cni.nih.gov/
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: ggafts@mail.nih.gov
            Tissue Procurement: Dr. Stefan Hansson
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) with help
            and advice from Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NCAM355 row: 5 column: 20
            High quality sequence stop: 578.
            Location/Qualifiers
             1..507
              /organism="Homo sapiens"
              /mol_type="mRNA"
FEATURES             source

```

```

/db_xref="taxon:9606"
/clone="IMAGE:3033474"
/tissue_type="pre-ecampic placenta"
/lab_host="CH1CB TotA"
/clone_lib="NH_MGC_148"
/note="Organ: placenta; Vectors: pBluescript1.9; Site: 1; all-XhoI; Site_2: BamI; Library is oligo dT primed and directionally cloned using primer: 5'-TTTTTTTCTTTTCTTCTT-3', size-selected for average insert size 2.3 kb and normalized to RCT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carroll, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH MGC library."
BASE COUNT      216 a      192 c      198 g      201 t
ORIGIN

Alignment Scores:
Pred. No.:      8,466 72      Length:      807
Score:          318.00      Matches:      175
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      100.00%      Indels:      0
DB:              14          Gaps:      0

US:10 029 117 2 (L 178; X CB992984 (L 807))

QY 1 MetAlaIaProLeuGlyGlyMetProSerGlyGlnProGlyProProGlyAlaPro 20
Db 11111111111111111111111111111111111111111111111111111
QY 13 ATGGCGGGCTCCACTAGGCGGTATGTTTCTTGGTAGCCATGGGTTCCGCTTCAGGCGCGG 92
Db 11111111111111111111111111111111111111111111111111111
QY 21 ProGlyLeuProGlyGlnLaserLeuLeuGlnAlaProGlyAlaProArgProSer 40
Db 11111111111111111111111111111111111111111111111111111
QY 43 CGGCGGCTTCGGGCGCAAGCTTCGCTTCCTTCAGCACTTCAGGCGCTTATAGACCTTC 152
Db 11111111111111111111111111111111111111111111111111111
QY 41 SerSerThrLeuValAspGlyLeuGlySerSerProGlyAlaCysPheAlaSerLeuVal 60
Db 11111111111111111111111111111111111111111111111111111
QY 153 AGCACTATCTTGGCGACGAGTTGGAGTCATCTTTCCAGGCGCTTCCTTCATCTTCGGT 212
Db 11111111111111111111111111111111111111111111111111111
QY 61 SerGlnAspTyrValAsnGlyThrAspGlnGlnGlnLeuArgThrGlyValAspGlyCys 80
Db 11111111111111111111111111111111111111111111111111111
QY 213 AATGAGGACTATGTCACTGGCACCGATCAGAGAAATATGACACTGCTGTATACATG 272
Db 11111111111111111111111111111111111111111111111111111
QY 81 LeuGlnLysProLeuAspIleLeuAArgGlnThrGlyCysPheLeuLeuLysArgLeu 300
Db 11111111111111111111111111111111111111111111111111111
QY 273 ATCCAGCAAGTTCTGGGATATGCCAAGACAGACAAATGTTTCTTCAGAAAGAAGT 332
Db 11111111111111111111111111111111111111111111111111111
QY 101 GlnLeuSerValGlnLysProGluGlnValIleLysGlnAspValSerGlnLeuArgSer 420
Db 11111111111111111111111111111111111111111111111111111
QY 333 CAGTCTATGTCTCCAGAACCCAGACAGTATCTAAAGAGAGCTGTCTATACTAAGGAAT 392
Db 11111111111111111111111111111111111111111111111111111
QY 121 GluLeuGlyArgLysAspAlaIleuValGlnLysPheThrLysLeuArgHisTrpGln 440
Db 11111111111111111111111111111111111111111111111111111
QY 393 GAATTACAGCGCAAGATGCCACTAGTCCAGAGCACTTACAGCAAGCTCAGGCAATTGGTAG 482
Db 11111111111111111111111111111111111111111111111111111
QY 141 GlnValLeuGluAspIleAsnValGlnHisLysLysProAlaAspIleArgGlnGlySer 460
Db 11111111111111111111111111111111111111111111111111111
QY 453 CAGGTGCTCGAGGAGCATCAAGCTGACACCAAAAGATGCTGACATTCCTCAGGCGTCC 512
Db 11111111111111111111111111111111111111111111111111111
QY 161 LeuAlaTyrLeuGluGlnAlaSerAlaAsnIleProAlaProLeuLysSerThr 538
Db 11111111111111111111111111111111111111111111111111111
QY 513 TGGTCTACCTGGACGACGATCTGCCAATATCTCTGACCTCTGAAGCTCAAGC 566
Db 11111111111111111111111111111111111111111111111111111

RESULT 5
BG141322
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
831 bp mRNA linear EST 01-FEB-2001
RG141322
NG5 HUVEC cDNA Library Homo sapiens cDNA, mRNA sequence.
BG141322
RG141322.1 GI:12643198
EST.

```

```

Db 498 CAGGTGCTGGAGACATCAACGTGACGACACAAAGCCGCGACATCCCTCAGGGCTCC 557
QY 161 LeuAlaTyrLeuCluGlnAlaSerAlaAlaSerAlaAlaProLeuLysProThr 178
Db 558 TTGGCTACTGAGGAGGACATCTGCCAAGATCCCTGACCTTCAAGCCAAACG 61:

RESULT 6
LOCUS BU158756 979 bp mRNA linear EST C4 SEP-2002
DEFINITION AGENCOURT_7844789 NIH_MGC 92 Homo sapiens cDNA clone IMAGE:544243
5', mRNA sequence.
ACCESSION BU158756
VERSION BU158756.1 GI:12572644
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (bases 1 to 879)
REFERENCE NIH_MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgaabs-fsmail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: X3T clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL3286 row: 1 column: 12
High quality sequence stop: 757.
FEATURES
source location/Qualifiers
1..879
/organism="Homo sapiens"
/mo_type="mRNA"
/db_xref="taxon:9606"
/c Gene=UAGS:604243"
/tissue type="embryonal carcinoma, cell line"
/lab host="DH10B (phage resistant)"
/clone lib="NIH_MGC_92"
/notes="Organ: testis; Vector: pCMV-SPORT6; Site: 1: Not2;
Site: 2: Salt; Cloned unidirectionally; oligo-dT primed;
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 217 a 192 c 211 g 259 t
ORIGIN
Alignment Scores:
Pred. No.: 9,488-71 Length: 888
Score: 918.00 Matches: 178
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0

US-10-029-137-2 (1-178) x BU158756 (1-879)
QY 1 MetAlaAlaProLeuGlyGlyMetPheSerGlyGlnProProGlyProGlnAlaPro 20
Db 6 ATGGCGCTCCACTAGGGGATATGTTTCTGGGACGACGACGCTCCCTCAGGCCCGG 65
QY 21 ProGlyLeuProGlyGlnAlaSerLeuSerGlnAlaAlaProGlyAlaProGlyPro 40
Db 66 CCGGGCTTCGGGCCCAAGCTTCGCTTCCTCAGCGACGACGACGCTCCTAGACCTCC 125
QY 41 SerSerThrLeuValAspGluLeuSerGlnAlaCysPheAlaSerLeuVal 60
Db 126 AGCAGTACTTGGTGGACGAGTTGGATCATCTTCAGGCTTGCCTTGCATCTCTGGTG 185
QY 61 SerGlnAspTyrValAsnGlyTyrAspGlnGluLeuAlaGlnThrGlyValAspGlnCys 80

```

```

Db 186 AGTCGAGACTATGTCAATGGCACCCATCAGAGAGAAATTCGAACCGGTGTGATCACTGT 245
QY 81 11eGlnCysPheLeuAspLeuAlaArgGlnThrGluCysPhePheLeuGlnCysArgLeu 100
Db 246 ATCCAGAGATCTCTGATATTGCAAGACAGACAGAGAAATGTTTCTTTTACAAAAAGATTG 305
QY 101 GlnLeuSerValGlnCysProGluGlnValIleLysGluAspValSerGluLeuArgAsn 120
Db 306 CAGTTATCTGTCAGAAACCCAGGCAAGTTATCAAAAGAGGATGTGTCAAGAACTAAGAAAT 365
QY 121 GluLeuGlnArgCysAspAlaLeuValGlnCysHisLeuThrLysLeuAlaGlnHisTrpGln 140
Db 366 GAATTACAGCGGAAGATGCACTAGTCCAGAGAGCACTTGACAAAGCTGAGSCATTGGCAG 425
QY 141 GlnValLeuGluAspLeuAsnValGlnHisLysLysProAlaAspLeuProGlnGlySer 160
Db 426 CAGGTCTGGAGGACATCAACGTGACGACACAAAGCCGCGACATCCCTCAGGGCTCC 495
QY 161 LeuAlaTyrLeuGluGlnAlaSerAlaAlaSerAlaAlaProLeuLysProThr 178
Db 486 TTGGCTACTGAGGAGGACATCTGCCAAGATCCCTGACCTTCAAGCCAAACG 539

RESULT 7
LOCUS CD389145 889 bp mRNA linear EST 30 MAY-2003
DEFINITION AGENCOURT_14305122 NIH_MGC 173 Homo sapiens cDNA 5', mRNA sequence.
ACCESSION CD389145
VERSION CD389145.1 GI:11227741
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (bases 1 to 889)
REFERENCE NIH_MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished
JOURNAL Contact: Daniela S. Gerhard, Ph.D.
COMMENT Office of Cancer Genomics
National Cancer Institute / NIH
Bldg 31 Rm10A07 Bethesda, MD 20892
Email: cgaabs-fsmail.nih.gov
Tissue Procurement: Dr. Jamie Thompson, University of WI
CDNA Library Preparation: Gena Zastrow-Hayes
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NBRX70 row: 1 column: 06
High quality sequence start: 28
High quality sequence stop: 747.
FEATURES
source location/Qualifiers
1..889
/organism="Homo sapiens"
/mo_type="mRNA"
/db_xref="taxon:9606"
/tissue type="trophoblast"
/lab host="DH10B Tona"
/clone lib="NIH_MGC_173"
/notes="Vector: pDONR201; Site: 1: attP2; Site: 2: attP1;
L19R-PRIME - oligo dT; METHOD - full-length enriched;
LIBR PROVIDER - Bradford"
BASE COUNT 223 a 198 c 215 g 252 t
ORIGIN
Alignment Scores:
Pred. No.: 9,488-71 Length: 888
Score: 918.00 Matches: 178
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0

```

```

DB:
US-10 029 137-2 (1-178) x CD389145 (1-888)
Gaps: 0

QY 1 MetAlaAlaProLeuGlyGlyMetPheSerGlyGlnProProGlyProGlnAlaPro 20
|||||
DB 44 ATGGCGCTCCACTAGGGGTATGTTTTCGGGACGCCACCGCTCCCTCAGGCCCG 103
|||||

QY 21 ProGlyLeuProGlyGlnAlaSerLeuGlnAlaAlaProGlyAlaProGProSer 40
|||||
DB 104 CCGGCGCTTCGGGCGAAGCTTCGCTCTTCAGGCGCTCCAGCGCTCTAGACCTTC 163
|||||

QY 41 SerSerThrLeuValAspGluLeuGluSerSerPheGluAlaCysPheAlaSerLeuVal 60
|||||
DB 164 ACCAGTACTTGTGGGACGAGTTGGAGTCACTTTTCAGGCTTGCCTTCATCTCTCGG 223
|||||

QY 61 SerGlnAspTyrValAspGlyThrAspGlnGluGluLeuArgThrGlyValAspGlnCys 80
|||||
DB 224 AGTATGAGACATGTCAATGGACCGATCAGGAGAGATTCAGACCGCTCTGATCGTCT 283
|||||

QY 81 TiedGlnLysPheLeuAspPheAlaAlaArgGlnThrGluCysPhePheLeuGlnLysAlaGlu 100
|||||
DB 284 ATCCAGAGTTCTGGGATATTGCAAGACACAGAGATGTTTTCACAAAAGATTCG 343
|||||

QY 101 GlnLeuSerValGlnLysProGluGlnValLleLysGluAspValSerGlnLeuArgAsn 120
|||||
DB 344 CAGTATCTGTGCCAGAACGAGCAAGTATTAAGAGGATGTGTCAAACTAAGGAAT 403
|||||

QY 121 GluLeuGlnArgLysAspAlaLeuValGlnLysHisLeuThrLysLeuArgHisLysPGLin 140
|||||
DB 404 GAATACAGCGGGAAGATGCACTAGTCCAGAGACATTCAGCAAGCTGAGCGCATTCG 463
|||||

QY 141 GlnValLeuGluAspIleAsnValGlnHisLysPheProAlaAspIleProGlnGlySer 160
|||||
DB 464 CAGTGTGTGAGGACATCAAGCTGCAGCAGAAAAGCCCGCAGACATCTATCAGGCTTC 523
|||||

QY 161 LeuAlaTyrLeuGluGlnAlaSerAlaAsnIleProAlaProLeuLysPheThr 178
|||||
DB 524 TTGGCTACTTGGGACGAGGCTCTGCCAATATCTGTGACTCTGAAGCGCAAGC 577
|||||

RESULT 8
BX437420 1019 bp mRNA linear EST 15-NOV-2003
LOCUS BX437420 Homo sapiens THYMUS H2O sapiens cDNA clone CS0CAP006Y020
DEFINITION 5-PRIME mRNA sequence.
ACCESSION BX437420
VERSION BX437420.1 GI:30779571
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization.
JOURNAL Unpublished
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 2813.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgl-bin/cluster.cgi?seq=CS0CAP006Y020&cluster=2813.f. Contact :
Feng Liang Email: fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0CAP006Y020
FEATURES
Location/Qualifiers
1..1019
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

/clone="CS0CAP006Y020"
/tissue_type="THYMUS"
/clone_lib="Homo sapiens THYMUS"
/notes="vector: pMSPORT_6; 1st strand cDNA was printed
with a NotI oligonucleotide primer. Five prime end enriched,
double strand cDNA was digested with NotI and cloned into
the NotI and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
BASE COUNT 258 a 255 g 265 t 18 others
CRIG:IN

Alignment Scores:
Pred. NO.: 1028 70 Length: 1019
Score: 818.00 Matches: 178
Percent Similarity: 100.000 Conserved: 0
Best Local Similarity: 100.000 Mismatches: 0
Query Matches: 100.000 Gaps: 0
DB: 100.000
DB: 100.000
US-10 029-137-2 (1-178) x BX437420 (1-1019)
QY 1 MetAlaAlaProLeuGlyGlyMetPheSerGlyGlnProProGlyProGlnAlaPro 20
|||||
DB 92 ATGGCGCTCCACTAGGGGTATGTTTTCGGGACGCCACCGCTCCCTCAGGCCCG 191
|||||

QY 81 ProGlyLeuProGlyGlnAlaSerLeuGlnAlaAlaProGlyAlaProGProSer 40
|||||
DB 192 CCGGCGCTTCGGGCGAAGCTTCGCTCTTCAGGCGCTCCAGCGCTCTAGACCTTC 211
|||||

QY 41 SerSerThrLeuValAspGluLeuGluSerSerPheGluAlaCysPheAlaSerLeuVal 60
|||||
DB 212 ACCAGTACTTGTGGGACGAGTTGGAGTCACTTTTCAGGCTTGCCTTCATCTCTCG 271
|||||

QY 61 SerGlnAspTyrValAspGlyThrAspGlnGluGluLeuArgThrGlyValAspGlnCys 80
|||||
DB 272 AGTATGAGACATGTCAATGGACCGATCAGGAGAGATTCAGACCGCTCTGATCGTCT 331
|||||

QY 81 TiedGlnLysPheLeuAspPheAlaAlaArgGlnThrGluCysPhePheLeuGlnLysAlaGlu 100
|||||
DB 332 CAGTATCTGTGCCAGAACGAGCAAGTATTAAGAGGATGTGTCAAACTAAGGAAT 391
|||||

QY 101 GlnLeuSerValGlnLysProGluGlnValLleLysGluAspValSerGlnLeuArgAsn 120
|||||
DB 392 CAGTATCTGTGCCAGAACGAGCAAGTATTAAGAGGATGTGTCAAACTAAGGAAT 451
|||||

QY 121 GluLeuGlnArgLysAspAlaLeuValGlnLysHisLeuThrLysLeuArgHisLysPGLin 140
|||||
DB 452 GAATACAGCGGGAAGATGCACTAGTCCAGAGACATTCAGCAAGCTGAGCGCATTCG 511
|||||

QY 141 GlnValLeuGluAspIleAsnValGlnHisLysPheProAlaAspIleProGlnGlySer 160
|||||
DB 512 CAGTGTGTGAGGACATCAAGCTGCAGCAGAAAAGCCCGCAGACATCTATCAGGCTTC 571
|||||

QY 161 LeuAlaTyrLeuGluGlnAlaSerAlaAsnIleProAlaProLeuLysPheThr 178
|||||
DB 572 TTGGCTACTTGGGACGAGGCTCTGCCAATATCTGTGACTCTGAAGCGCAAGC 628
|||||

RESULT 9
AF18059 1077 bp mRNA linear HTC 02 NOV-1991
LOCUS AF18059 Homo sapiens FKSG20 (FKSG20) mRNA, complete cds.
DEFINITION AF18059
ACCESSION AF18059
VERSION AF18059.1 GI:1658815
KEYWORDS HTC
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1077)
AUTHORS Wang,Y. G. and Gong,L.
TITLE Cloning of FKSG20, a novel gene expressed in Adrenal gland tumor
tissues
JOURNAL Unpublished

```

```

RESULT 10
BX376389
LOCUS
DEFINITION
BX376389 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
CDNA clone CS00C27YE12 5'-PRIME, mRNA sequence.
BX376389
BX376388.1 G133344657
EST
BX376388.1 G133344657
Homo sapiens (human)
Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases) to 1173
Li W.B., Gruber C., Cessari C. and Polayes D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
Ap 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 2813.f for
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS00C27BC06QPlcluster=2813.f. Contact :
Peng Liang Email : liang@life-techn.com URL :
http://fulllength.invitrogen.com/ invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS00C27BC06QPl
Location/Qualifiers
1..1173
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS00C27YE12"
/clone_type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/notes="First strand cDNA was primed with a NotI-oligo(dT)
primer. Five copies and enriched double-strand cDNA was
digested with NotI and cloned into the NotI and EcoRV
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT 299 a 240 c 269 g 103 t 62 others
ORIGIN
Alignment Scores:
P-Id. No. 1 31670 Length: 1173
Score 916.20 Matches: 179
Percent Similarity: 100.000 Conservative: 0
Best Local Similarity: 100.000 W.smatches: 0
Query Match: 100.000 Indels: 0
DB: 13 Gaps: 0
US 10-029-137-2 (1178) x BX376388 (11173)
QY 1 MetAlaLaAProLeSGlyGlyMetPreSerGlyGlnProProGlyProGlnAlaPro 20
Dh 64 ATGGGAGGATCCACTAGGGGGGATGTTCTGGGGAGCCACCAGCGGTCCTCGAGCCCG 123
QY 21 ProGlyGluProGlyGlnAlaSerLeuGlnAlaAlaProGlyValaProArgProSer 40
Db 124 CCGGAGCTTCGGGSCCAAGCTTCGCTTCTTCAGGCAGCTCCAGGCGCTCCTAGACCTTC 183
QY 41 SerSerThrLeuValAsnGluLeuGluSerSerPheGluAlaCysPheAlaSerLeuVal 60
Db 184 AGCAGTACTTTGGTGGACGAGTGGAGTCATCTTCAGGGCTTGGTTTCATCTCTGGTG 243
QY 61 SerGlnAspTyrValAsnGlyThrAspGlnGluGluLeuArgThrGlyValAspGlnCys 80
Db 244 ACTCAGACTATGTCATGTCATGSCACCGATCAGAGAGAAATTCGAACCGGCTTCGATCAGT 303
QY 81 IleGlnCysPheLeuAspIleAlaArgGlnThrGlyCysPhePheLeuGlnCysArgLeu 100
Db 304 ATCCAGAAATTCGGATATTATCAACACACACACAAATGTTTTTCTTCACAAAAAGATTG 161

```



[illegible]









Db 262 MGR 265

# RESULT 2

US-10-020-079-10  
 : Sequence 10, Application US/10020079  
 : Patent No. 6579710  
 : GENERAL INFORMATION:  
 : APPLICANT: Turner, C. Alexander Jr.  
 : APPLICANT: Mathai, Brian  
 : APPLICANT: Friddle, Carl Johan  
 : TITLE OF INVENTION: No. 6579710el Human Kinases and Polynucleotides Encoding the Same  
 : FILE REFERENCE: JEX-0281-USA  
 : CURRENT APPLICATION NUMBER: US/10/020,079  
 : PRIOR FILING DATE: 2001-12-12  
 : PRIOR APPLICATION NUMBER: US 60/255,103  
 : PRIOR FILING DATE: 2000-12-12  
 : PRIOR APPLICATION NUMBER: US 60/269,422  
 : PRIOR FILING DATE: 2001-05-06  
 : NUMBER OF SEQ ID NOS: 40  
 : SOFTWARE: FastSeq for Windows Version 4.0  
 : SEQ ID NO 10  
 : LENGTH: 654  
 : TYPE: PRT  
 : ORGANISM: homo sapiens  
 : NAME/KEY: VARIANT  
 : LOCATION: (1) (1654)  
 : OTHER INFORMATION: Xaa = Any Amino Acid

US-10-029-079-10

Query Match 10.88; Score 96.5; DP 4; Length 654;  
 Best Local Similarity 23.48; Pred. No. 0.35;  
 Matches 57; Conservativ 25; Mismatches 63; Indels 99; Gaps 11;

QY 2 AAPLGWFSQQF-----PQPQAPGLPGQASL-----LOAAPGAP 37  
 |||  
 Db 43 AANAGS-SGSPNNA-SRPRPAPGCAAPRAASAAAGAAAGTQVQGLPAGPACT 100  
 |||  
 QY 18 RPS-SSSTJVDLESLSFACFASLSVQIV-----NGTQDEIRITGVQV 79  
 |||  
 Db 101 PPSWRWQCLAAALQDETNKSGGSEQADIPANVYVQKRAKVLKKGGSFGRIVEAMN 160  
 |||  
 QY 80 CIOKFLDIARQTEFFCKRLQLSVQKDEWIKEDUSEPNEIQRKD----- 126  
 |||  
 Db 161 LTRP--KVALKE-----SAQPKQVLEKVEAVK-KKQKQKHNCRFPGQORNEK 207  
 |||  
 QY 127 ---ALVKHRTKLRHWQ-----QVLEKINVCKR---KPADIP 157  
 |||  
 Db 208 PNYVWVQGRKJADLRSPGPTFTLTSLTRKQKQLSLRAHSWGFHRDIPKSNFA 267  
 |||  
 QY 154 QGSL 161  
 Db 268 MGR 271

# RESULT 3

US-10-020-079-4  
 : Sequence 4, Application US/10020079  
 : Patent No. 6579710  
 : GENERAL INFORMATION:  
 : APPLICANT: Turner, C. Alexander Jr.  
 : APPLICANT: Mathai, Brian  
 : APPLICANT: Friddle, Carl Johan  
 : TITLE OF INVENTION: No. 6579710el Human Kinases and Polynucleotides Encoding the Same  
 : FILE REFERENCE: JEX-0281-USA  
 : CURRENT APPLICATION NUMBER: US/10/020,079  
 : PRIOR FILING DATE: 2001-12-12  
 : PRIOR APPLICATION NUMBER: US 60/255,103  
 : PRIOR FILING DATE: 2000-12-12  
 : PRIOR APPLICATION NUMBER: US 60/289,422  
 : PRIOR FILING DATE: 2001-05-06  
 : NUMBER OF SEQ ID NOS: 40

: SOFTWARE: FastSeq for Windows Version 4.0  
 : SEQ ID NO 4  
 : LENGTH: 664  
 : TYPE: PRT  
 : ORGANISM: homo sapiens  
 : NAME/KEY: VARIANT  
 : LOCATION: (1) (1664)  
 : OTHER INFORMATION: Xaa = Any Amino Acid  
 : US-10-020-079-4

Query Match 10.88; Score 96.5; DP 4; Length 664;

Best Local Similarity 23.48; Pred. No. 0.35;

Matches 57; Conservativ 25; Mismatches 63; Indels 99; Gaps 11;

QY 2 AAPLGWFSQQF-----PQPQAPGLPGQASL-----LOAAPGAP 37  
 |||  
 Db 43 AANAGS-SGSPNNA-SRPRPAPGCAAPRAASAAAGAAAGTQVQGLPAGPACT 100  
 |||  
 QY 18 RPS-SSSTJVDLESLSFACFASLSVQIV-----NGTQDEIRITGVQV 79  
 |||  
 Db 101 PPSWRWQCLAAALQDETNKSGGSEQADIPANVYVQKRAKVLKKGGSFGRIVEAMN 160  
 |||  
 QY 80 CIOKFLDIARQTEFFCKRLQLSVQKDEWIKEDUSEPNEIQRKD----- 126  
 |||  
 Db 161 LTRP--KVALKE-----SAQPKQVLEKVEAVK-KKQKQKHNCRFPGQORNEK 207  
 |||  
 QY 127 ---ALVKHRTKLRHWQ-----QVLEKINVCKR---KPADIP 157  
 |||  
 Db 208 PNYVWVQGRKJADLRSPGPTFTLTSLTRKQKQLSLRAHSWGFHRDIPKSNFA 267  
 |||  
 QY 154 QGSL 161  
 Db 268 MGR 271

# RESULT 4

US-10-020-079-2  
 : Sequence 2, Application US/10020079  
 : Patent No. 6579710  
 : GENERAL INFORMATION:  
 : APPLICANT: Turner, C. Alexander Jr.  
 : APPLICANT: Mathai, Brian  
 : APPLICANT: Friddle, Carl Johan  
 : TITLE OF INVENTION: No. 6579710el Human Kinases and Polynucleotides Encoding the Same  
 : FILE REFERENCE: JEX-0281-USA  
 : CURRENT APPLICATION NUMBER: US/10/020,079  
 : PRIOR FILING DATE: 2001-12-12  
 : PRIOR APPLICATION NUMBER: US 60/255,103  
 : PRIOR FILING DATE: 2000-12-12  
 : PRIOR APPLICATION NUMBER: US 60/289,422  
 : PRIOR FILING DATE: 2001-05-06  
 : NUMBER OF SEQ ID NOS: 40  
 : SOFTWARE: FastSeq for Windows Version 4.0  
 : SEQ ID NO 2  
 : LENGTH: 664  
 : TYPE: PRT  
 : ORGANISM: homo sapiens  
 : NAME/KEY: VARIANT  
 : LOCATION: (1) (1670)  
 : OTHER INFORMATION: Xaa = Any Amino Acid  
 : US-10-020-079-2

Query Match 10.88; Score 96.5; DP 4; Length 670;

Best Local Similarity 23.48; Pred. No. 0.35;

Matches 57; Conservativ 25; Mismatches 63; Indels 99; Gaps 11;

QY 2 AAPLGWFSQQF-----PQPQAPGLPGQASL-----LOAAPGAP 37  
 |||  
 Db 43 AANAGS-SGSPNNA-SRPRPAPGCAAPRAASAAAGAAAGTQVQGLPAGPACT 100  
 |||  
 QY 36 RPS-----SSTJVDLESLSFACFASLSVQIV-----NGTQDEIRITGVQV 79

```
101 PPSGWMQCLAAALKDETNNSGGGQACILIPANYVVKKKVKKKGGGGEIYAVOL 160
102 CLOKFLDIARQTEFFLQRLCLSVCKEYVKEVSEIRNELQK 126
103 LTR--NVALKVE-----SAGQKQVLMKMEVAVLK KLOKQHVCFICGRNEK 207
104 ALVKHKLKLRHQ-----QVLELINVQHK-----KPADIP 157
105 FNVVWVQIQORNLADLRSCPRFTLTSLTLKSKLLESTEAHSVGLHRLDKPSNFA 267
106 QGSL 161
107 MGSL 261
```

```
Query Match 10.53; Score 96.5; DB 4; Length 895;
Best Local Similarity 23.43; Pred. No. 0.52;
Matches 57; Conservative 25; Mismatches 63; Indels 99; Gaps 11;

CY 2 AAPLGMVFSGGP-----FGPPQAPGGLPGQASL-----LQAAGCAP 37
DB 47 AAAGG--SSPNALSRPRAPAFAPDAPPRAAASAAAAAAGTGVDSPLRAGPADT 94
CY 38 RES-----SSTLVDELESSPEACFASLVSCYV-----NGTDGEETIGVDQ 79
DB 95 PPSGWMQCLAAALKDETNNSGGGQACILIPANYVVKKKVKKKGGGGEIYAVOL 154
CY 80 CLOKFLDIARQTEFFLQRLCLSVCKEYVKEVSEIRNELQK 126
DB 165 LTR--NVALKVE-----SAGQKQVLMKMEVAVLK KLOKQHVCFICGRNEK 201
CY 127 ALVKHKLKLRHQ-----QVLELINVQHK-----KPADIP 157
DB 202 FNVVWVQIQORNLADLRSCPRFTLTSLTLKSKLLESTEAHSVGLHRLDKPSNFA 261
CY 158 QGSL 161
DB 262 MGSL 261
```

```
Query Match 10.53; Score 96.5; DB 4; Length 895;
Best Local Similarity 23.43; Pred. No. 0.52;
Matches 57; Conservative 25; Mismatches 63; Indels 99; Gaps 11;

CY 2 AAPLGMVFSGGP-----FGPPQAPGGLPGQASL-----LQAAGCAP 37
DB 47 AAAGG--SSPNALSRPRAPAFAPDAPPRAAASAAAAAAGTGVDSPLRAGPADT 94
CY 38 RES-----SSTLVDELESSPEACFASLVSCYV-----NGTDGEETIGVDQ 79
DB 95 PPSGWMQCLAAALKDETNNSGGGQACILIPANYVVKKKVKKKGGGGEIYAVOL 154
CY 80 CLOKFLDIARQTEFFLQRLCLSVCKEYVKEVSEIRNELQK 126
DB 165 LTR--NVALKVE-----SAGQKQVLMKMEVAVLK KLOKQHVCFICGRNEK 201
CY 127 ALVKHKLKLRHQ-----QVLELINVQHK-----KPADIP 157
DB 202 FNVVWVQIQORNLADLRSCPRFTLTSLTLKSKLLESTEAHSVGLHRLDKPSNFA 261
CY 158 QGSL 161
DB 262 MGSL 261
```

```
Query Match 10.53; Score 96.5; DB 4; Length 895;
Best Local Similarity 23.43; Pred. No. 0.52;
Matches 57; Conservative 25; Mismatches 63; Indels 99; Gaps 11;

CY 2 AAPLGMVFSGGP-----FGPPQAPGGLPGQASL-----LQAAGCAP 37
DB 47 AAAGG--SSPNALSRPRAPAFAPDAPPRAAASAAAAAAGTGVDSPLRAGPADT 94
CY 38 RES-----SSTLVDELESSPEACFASLVSCYV-----NGTDGEETIGVDQ 79
DB 95 PPSGWMQCLAAALKDETNNSGGGQACILIPANYVVKKKVKKKGGGGEIYAVOL 154
CY 80 CLOKFLDIARQTEFFLQRLCLSVCKEYVKEVSEIRNELQK 126
DB 165 LTR--NVALKVE-----SAGQKQVLMKMEVAVLK KLOKQHVCFICGRNEK 201
CY 127 ALVKHKLKLRHQ-----QVLELINVQHK-----KPADIP 157
DB 202 FNVVWVQIQORNLADLRSCPRFTLTSLTLKSKLLESTEAHSVGLHRLDKPSNFA 261
CY 158 QGSL 161
DB 262 MGSL 261
```

```
GENERAL INFORMATION:
APPLICANT: Turner, C. Alexander Jr.
APPLICANT: Mathur, Brian
APPLICANT: Fridlie, Carl Johan
TITLE OF INVENTION: NO. 6579710e1 Human Kinases and Polynucleotides Encoding the
FILE REFERENCE: LEX-0281-USA
CURRENT APPLICATION NUMBER: US/10/020,079
CURRENT FILING DATE: 2001-12-12
PRIOR APPLICATION NUMBER: US 60/255,103
PRIOR FILING DATE: 2003-12-12
PRIOR APPLICATION NUMBER: US 60/289,422
PRIOR FILING DATE: 2001-05-08
NUMBER OF SEQ ID NOS: 40
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 18
LENGTH: 895
TYPE: PRT
ORGANISM: homo sapiens
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1).....(95)
OTHER INFORMATION: Xaa - Any Amino Acid
US-10-020-079-18
```

```
Query Match 10.53; Score 96.5; DB 4; Length 895;
Best Local Similarity 23.43; Pred. No. 0.52;
Matches 57; Conservative 25; Mismatches 63; Indels 99; Gaps 11;

CY 2 AAPLGMVFSGGP-----FGPPQAPGGLPGQASL-----LQAAGCAP 37
DB 47 AAAGG--SSPNALSRPRAPAFAPDAPPRAAASAAAAAAGTGVDSPLRAGPADT 100
CY 38 RES-----SSTLVDELESSPEACFASLVSCYV-----NGTDGEETIGVDQ 79
DB 101 PPSGWMQCLAAALKDETNNSGGGQACILIPANYVVKKKVKKKGGGGEIYAVOL 160
CY 80 CLOKFLDIARQTEFFLQRLCLSVCKEYVKEVSEIRNELQK 126
DB 161 LTR--NVALKVE-----SAGQKQVLMKMEVAVLK KLOKQHVCFICGRNEK 207
CY 127 ALVKHKLKLRHQ-----QVLELINVQHK-----KPADIP 157
DB 208 FNVVWVQIQORNLADLRSCPRFTLTSLTLKSKLLESTEAHSVGLHRLDKPSNFA 267
CY 158 QGSL 161
DB 268 MGSL 261
```

```
Query Match 10.53; Score 96.5; DB 4; Length 895;
Best Local Similarity 23.43; Pred. No. 0.52;
Matches 57; Conservative 25; Mismatches 63; Indels 99; Gaps 11;

CY 2 AAPLGMVFSGGP-----FGPPQAPGGLPGQASL-----LQAAGCAP 37
DB 47 AAAGG--SSPNALSRPRAPAFAPDAPPRAAASAAAAAAGTGVDSPLRAGPADT 100
CY 38 RES-----SSTLVDELESSPEACFASLVSCYV-----NGTDGEETIGVDQ 79
DB 101 PPSGWMQCLAAALKDETNNSGGGQACILIPANYVVKKKVKKKGGGGEIYAVOL 160
CY 80 CLOKFLDIARQTEFFLQRLCLSVCKEYVKEVSEIRNELQK 126
DB 161 LTR--NVALKVE-----SAGQKQVLMKMEVAVLK KLOKQHVCFICGRNEK 207
CY 127 ALVKHKLKLRHQ-----QVLELINVQHK-----KPADIP 157
DB 208 FNVVWVQIQORNLADLRSCPRFTLTSLTLKSKLLESTEAHSVGLHRLDKPSNFA 267
CY 158 QGSL 161
DB 268 MGSL 261
```

LOCATION: (1)....(951)  
OTHER INFORMATION: Xaa = Any Amino Acid  
US-10-029-079-36

Query Match 10.5%; Score 96.5; DB 4; Length 957;

Best Local Similarity 23.4%; Pred. No. 0.57;

Matches 57; Conservative 25; Mismatches 63; Indels 99; Gaps 11;

```

QY 2 AAPLGSMFSGQP-----PGPPQAFPGCPQCAASL-----LCNAPCAP 37
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 37 AAAGG--SGSNAA--SRPRAPAPGADAPPRAAASAAAAAAGTAGTCVQGPLRAGPACT 94
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 38 RPS-----SSTLVELESSFEACFASLSVDYV-----NGTQDEIRTVQDQ 79
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 95 PPSGWMQCLAAALKKDETNNYSGGGEQADILPANYVVKRWKVKLKGKGGGGEIYEAMOL 154
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 80 CIQKFLDIARQTECFEFLQKRLQLSVQKPEQVVKEDVSELRNELORKD----- 156
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 155 LTRR--NVALKVE-----SAQQPKVULKMEVAVLK-KLQCKDHVCFICGGGNEK 201
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 127 -- ALVQKHILTKLRHWQ-----OVLEDIRVCHK-----KPADIP 157
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 202 FNVVMVQGLGRNLADLRSSQPRGTFTLSTTLRLKQKQLESIEAHSVGFHLRDKKPSNFA 261
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 158 QGSL 161
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 262 MGRLL 265
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

#### RESULT 8

US-10-020-079 34

Sequence 34, Application US/10020079

Patent No. 6579710

GENERAL INFORMATION:

APPLICANT: Turner, C. Alexander Jr.

APPLICANT: Mathur, Bilan

APPLICANT: Fridele, Carl Johan

TITLE OF INVENTION: No. 6579710e: Human Kinases and Polynucleotides Encoding the Same

FILE REFERENCE: LEX 0281-USA

CURRENT APPLICATION NUMBER: US/10/020,079

PRIOR FILING DATE: 2001-12-12

PRIOR APPLICATION NUMBER: US 60/255,101

PRIOR FILING DATE: 2000-12-12

PRIOR APPLICATION NUMBER: US 60/289,422

PRIOR FILING DATE: 2001-05-08

NUMBER OF SEQ ID NOS: 40

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 34

LENGTH: 957

TYPE: PRT

ORGANISM: homo sapiens

FEATURE:

NAME/KEY: VARIANT

LOCATION: (1)....(957)

OTHER INFORMATION: Xaa = Any Amino Acid

US-10-020 079-34

Query Match 10.5%; Score 96.5; DB 4; Length 957;

Best Local Similarity 23.4%; Pred. No. 0.57;

Matches 57; Conservative 25; Mismatches 63; Indels 99; Gaps 11;

```

QY 2 AAPLGSMFSGQP-----PGPPQAFPGCPQCAASL-----LCNAPCAP 37
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 43 AAAGG--SGSNAA--SRPRAPAPGADAPPRAAASAAAAAAGTAGTCVQGPLRAGPACT 100
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 38 RPS-----SSTLVELESSFEACFASLSVDYV-----NGTQDEIRTVQDQ 79
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 101 PPSGWMQCLAAALKKDETNNYSGGGEQADILPANYVVKRWKVKLKGKGGGGEIYEAMOL 160
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 80 CIQKFLDIARQTECFEFLQKRLQLSVQKPEQVVKEDVSELRNELORKD----- 126
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 161 LTRR--NVALKVE-----SAQQPKVULKMEVAVLK-KLQCKDHVCFICGGGNEK 207
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY 127 -- ALVQKHILTKLRHWQ-----OVLEDIRVCHK-----KPADIP 157
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 208 FNVVMVQGLGRNLADLRSSQPRGTFTLSTTLRLKQKQLESIEAHSVGFHLRDKKPSNFA 267
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 158 QGSL 161
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 268 MGRLL 271
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

#### RESULT 9

US-10-020-079-26

Sequence 26, Application US/10020079

Patent No. 6579710

GENERAL INFORMATION:

APPLICANT: Turner, C. Alexander Jr.

APPLICANT: Mathur, Bilan

APPLICANT: Fridele, Carl Johan

TITLE OF INVENTION: No. 6579710e: Human Kinases and Polynucleotides Encoding the Same

FILE REFERENCE: LEX 0281-USA

CURRENT APPLICATION NUMBER: US/10/020,079

PRIOR FILING DATE: 2001-12-12

PRIOR APPLICATION NUMBER: US 60/255,101

PRIOR FILING DATE: 2000-12-12

PRIOR APPLICATION NUMBER: US 60/289,422

PRIOR FILING DATE: 2001-05-08

NUMBER OF SEQ ID NOS: 40

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 28

LENGTH: 974

TYPE: PRT

ORGANISM: homo sapiens

FEATURE:

NAME/KEY: VARIANT

LOCATION: (1)....(974)

OTHER INFORMATION: Xaa = Any Amino Acid

US-10-020-079-28

Query Match 10.5%; Score 96.5; DB 4; Length 976;

Best Local Similarity 23.4%; Pred. No. 0.57;

Matches 57; Conservative 25; Mismatches 63; Indels 99; Gaps 11;

```

QY 2 AAPLGSMFSGQP-----PGPPQAFPGCPQCAASL-----LCNAPCAP 37
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 37 AAAGG--SGSNAA--SRPRAPAPGADAPPRAAASAAAAAAGTAGTCVQGPLRAGPACT 94
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 38 RPS-----SSTLVELESSFEACFASLSVDYV-----NGTQDEIRTVQDQ 79
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 95 PPSGWMQCLAAALKKDETNNYSGGGEQADILPANYVVKRWKVKLKGKGGGGEIYEAMOL 154
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 80 CIQKFLDIARQTECFEFLQKRLQLSVQKPEQVVKEDVSELRNELORKD----- 156
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 155 LTRR--NVALKVE-----SAQQPKVULKMEVAVLK-KLQCKDHVCFICGGGNEK 201
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 127 -- ALVQKHILTKLRHWQ-----OVLEDIRVCHK-----KPADIP 157
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 202 FNVVMVQGLGRNLADLRSSQPRGTFTLSTTLRLKQKQLESIEAHSVGFHLRDKKPSNFA 261
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 158 QGSL 161
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 262 MGRLL 265
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

#### RESULT 10

US-10-020-079-26

Sequence 26, Application US/10020079

Patent No. 6579710

GENERAL INFORMATION:

APPLICANT: Turner, C. Alexander Jr.

APPLICANT: Mathur, Bilan

APPLICANT: Fridele, Carl Johan

TITLE OF INVENTION: No. 6579710e: Human Kinases and Polynucleotides Encoding the Same

FILE REFERENCE: LEX 0281-USA

CURRENT APPLICATION NUMBER: US/10/020,079



? CURRENT FILING DATE: 2001-12-12  
 ? PRIOR APPLICATION NUMBER: US 60/255,103  
 ? PRIOR FILING DATE: 2000-12-12  
 ? PRIOR APPLICATION NUMBER: US 60/289,422  
 ? PRIOR FILING DATE: 2001-05-08  
 ? NUMBER OF SEQ ID NOS: 40  
 ? SOFTWARE: FASTSEQ for Windows Version 4.0  
 ? SEQ ID NO 26  
 ? LENGTH: 982  
 ? TYPE: PRT  
 ? ORGANISM: Homo sapiens  
 ? FEATURE:  
 ? NAME/KEY: VARIANT  
 ? LOCATION: (1)...(982)  
 ? OTHER INFORMATION: Xaa = Any Amino Acid  
 US-10-029-073-26

Query Match 10.53; Score 96.5; DB 4; Length 982;  
 Best Local Similarity 23.4%; Pident No. 0.59;  
 Matches 57; Conservative 25; Mismatches 63; Indels 99; Gaps 11;  
 QY 2 AAPGMPFSGP... --- PGPQAPFGLPGQASL... --- LQAPGAP 37  
 DB 43 ALAAGG--SSSPNALSRRPAPAPGADAPFAAASAAAAAATGVTGTPLAGPAT 100  
 QY 1P RFS...---SSTANFLESSPEAFSLVQCYV... --- NOTGEEFTGVIC 79  
 DB 101 PESTAWMOCLAAAGCTNMSGGBOACILPANYVYKDPKAVKIKIGGGPGEYVAVEL 160  
 QY 40 CCKKELDIARQTECFELQKRLSVKQTEQVKEVSVIRNELQKQ... --- 126  
 DB 161 LTRK--KVALKVE...--- SAQKQVLMFVAVLK-KLQKDHVCFPGCGRNEK 207  
 QY 127 -...--ALVKKLTKLEHW... --- ---QVLEDINVQK...--- KPADIP 157  
 DB 208 FNVVVCQICGRNLARLRCQFQCTFLSTPLGKQVLESLEHSHVFLHRLKPSNFA 267  
 QY 119 QSSV 161  
 DB 148 MRL 171

RESULT 1)  
 US 08 348 518C-4  
 ? Sequence 4, Application US'08348518C  
 ? Patent No. 6022740  
 ? GENERAL INFORMATION:  
 ? APPLICANT: SUDOL, MARCUS  
 ? APPLICANT: PERRY, BORK  
 ? APPLICANT: HENRY, CHEN  
 ? TITLE OF INVENTION: A SH2 DOMAIN ASSOCIATED PROTEIN, A  
 ? TITLE OF INVENTION: SIGNALING DOMAIN THEREIN, NUCLEIC ACIDS ENCODING THE  
 ? TITLE OF INVENTION: PROTEIN AND THE DOMAIN, AND DIAGNOSTIC AND THERAPEUTIC USES  
 ? NUMBER OF SEQUENCES: 25  
 ? CORRESPONDENCE ADDRESS:  
 ? ADDRESSEE: Klauber & Jackson  
 ? STREET: 411 Hackensack Avenue  
 ? CITY: Hackensack  
 ? STATE: New Jersey  
 ? COUNTRY: USA  
 ? ZIP: 07601  
 ? COMPUTER READABLE FORM:  
 ? MEDIUM TYPE: Floppy disk  
 ? COMPUTER: IBM PC compatible  
 ? OPERATING SYSTEM: PC-DOS/MS-DOS  
 ? SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ? CURRENT APPLICATION DATA:  
 ? APPLICATION NUMBER: US/08/348,518C  
 ? FILING DATE: 01-DEC-1994  
 ? CLASSIFICATION: 514  
 ? ATTORNEY/AGENT INFORMATION:  
 ? NAME: Jackson Esq., David A.

? REGISTRATION NUMBER: 26,742  
 ? REFERENCE/DOCKET NUMBER: 600-1-101;  
 ? TELEPHONE: 201 487-5800  
 ? TELEFAX: 201 343-1684  
 ? TELEX: 133521  
 ? INFORMATION FOR SEQ ID NO: 4:  
 ? SEQUENCE CHARACTERISTICS:  
 ? LENGTH: 454 amino acids  
 ? TYPE: amino acid  
 ? TOPOLOGY: linear  
 ? MOLECULE TYPE: protein  
 ? US-38 348-518C-4

Query Match 10.53; Score 95; DB 3; Length 454;  
 Best Local Similarity 25.0%; Pident No. 0.3;  
 Matches 46; Conservative 21; Mismatches 59; Indels 58; Gaps 10;  
 QY 3 APGGMFSGP...---CAPSLRQ...---ASLIQAPGAPRFSSTLV... DEKSSFEA 54  
 DB 12 APQG...---GQPSQFQSGGPGGQAPAPATQAAPCAP PAGCIVHVRGDSSTILEA 67  
 QY 55 CASLVSCYVNGTJOEEELRTSVQCIQKFLDIARQTECFELQKRLSVKQTEQVKEVSVIRNELQKQ... 114  
 DB 68 LFNAMVNTAN...--- --VPQTPMRL...--- ---RKLPSEFPKQPEP-RSH 104  
 QY 115 VSELRKELQKCALVKKLTKLKHQV...LEQVQKHPAD:FGSLAYLEQASANIDAP 174  
 DB 105 SPQASTAGTAGALTECHVR... --- ---AHSSPASICLQAVS... ---POT 141  
 QY 175 LKPT 175  
 DB 142 LTP 145

RESULT 12  
 US-08-476-509B-4  
 ? Sequence 4, Application US/08476509B  
 ? Patent No. 6034212  
 ? GENERAL INFORMATION:  
 ? APPLICANT: SUDOL, MARCUS  
 ? APPLICANT: PERRY, BORK  
 ? APPLICANT: HENRY, CHEN  
 ? TITLE OF INVENTION: A SH2 DOMAIN ASSOCIATED PROTEIN, A  
 ? TITLE OF INVENTION: SIGNALING DOMAIN THEREIN, NUCLEIC ACIDS ENCODING THE  
 ? TITLE OF INVENTION: PROTEIN AND THE DOMAIN, AND DIAGNOSTIC AND THERAPEUTIC USES  
 ? NUMBER OF SEQUENCES: 50  
 ? CORRESPONDENCE ADDRESS:  
 ? ADDRESSEE: Klauber & Jackson  
 ? STREET: 411 Hackensack Avenue  
 ? CITY: Hackensack  
 ? STATE: New Jersey  
 ? COUNTRY: USA  
 ? ZIP: 07601  
 ? COMPUTER READABLE FORM:  
 ? MEDIUM TYPE: Floppy disk  
 ? COMPUTER: IBM PC compatible  
 ? OPERATING SYSTEM: PC-DOS/MS-DOS  
 ? SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ? CURRENT APPLICATION DATA:  
 ? APPLICATION NUMBER: US/08/476,509B  
 ? FILING DATE: 01-DEC-1994  
 ? CLASSIFICATION:  
 ? ATTORNEY/AGENT INFORMATION:  
 ? NAME: Jackson Esq., David A.  
 ? REGISTRATION NUMBER: 26,742  
 ? REFERENCE/DOCKET NUMBER: 600-1-101 CIP  
 ? TELECOMMUNICATION INFORMATION:  
 ? TELEPHONE: 201 487-5800  
 ? TELEFAX: 201 343-1684  
 ? TELEX: 133521  
 ? INFORMATION FOR SEQ ID NO: 4:



24 422 ANSPGOLLTPALLERRI LSDVYQGLPHANSACL BELKSSYHE-----YRYFET 270  
CY 70 QBEIRTVQCTQKFLDIARQTECFPLQXR --LOLSYCKPEQVIXED-- --- ---- 114  
26 271 QHQ---SVFQCLSKFUCASPEIMNVHTAVPSLQHLKALLNEVILDECHKEKIVCTWETO 327  
CY 115 --VSELRNELOPKUVALVOKHILTKLRH-WOQVLEQINVQHKHAG 155  
28 328 ELVSEAYPILEQKIKULQPHVVASNNWEPRISQYCKIKLRNTQ 371

Search completed: October 21, 2003, 16:11:17  
Job time: 1.47 secs